Fuzzy Approach to Ecological Data Analysis

ARKADIUSZ SALSKI

Department of Computer Science

University of Kiel
Olshausenstrasse 40, D-24098 Kiel
Germany
http://www.is.informatik.uni-kiel.de/~asa

Abstract: The paper focuses on the fuzzy extensions of two data analysis methods, namely the fuzzy cluster analysis and the fuzzy interpolation of spatial data (the so-called fuzzy kriging) and their suitability for ecological applications. Both extensions utilize exact (crisp) measurement data as well as imprecise data defined as fuzzy numbers or fuzzy vectors. Fuzzy clustering of fuzzy data (conical fuzzy vectors) and fuzzy kriging methods are useful tools with a potential of utilizing vast pieces of information available as inexact or uncertain data. We have to deal with such data in many ecological applications. The paper presents an example of the analysis of ecological data and two fuzzy data analysis systems developed at the University of Kiel.

Key-Words: fuzzy clustering, conical fuzzy vector, fuzzy classification, fuzzy kriging, fuzzy regionalization, ecological application

1 Uncertainty and Heterogeneity of Ecological Data

Besides the usual problem of searching for effective methods for data analysis there is a further difficulty with the analysis of ecological data caused by the uncertainty and heterogeneity of these data. Ecologists collect and evaluate data from all possible data sources, sources of objective (mostly quantitative) data, like measurements and simulation results, and sources of subjective (often only qualitative) information, like subjective estimations obtained from an expert. Not all ecological parameters are measurable, for example the number and biomass of fish in a particular lake.

The uncertainty and heterogeneity problems in ecological data analysis result from:

- presence of random variables,
- incomplete or inaccurate data,
- approximate estimations instead of measurements,
- incomparability of data (varying measurement or observation conditions),
- imprecise qualitative instead of quantitative data,
- subjectivity of the information obtained from expert, and
- heterogeneous data sources.

The requirements for the methods of ecological data analysis arise from these properties. Thus, special methods should be used to handle imprecision, uncertainty and heterogeneity of environmental data. Conventional data analysis methods based on Boolean logic often ignore the continuous nature of ecological parameters and the uncertainty of ecological data. That can result in misclassification or misinterpretation of the data structure. Fuzzy representation and interpretation of data structure is a very natural and intuitively plausible way to formulate and to solve some uncertainty problems in ecological data analysis but in a different way as the probabilistic approach. Ecological data or classes of ecological objects can be defined as fuzzy sets with no sharply formed boundaries, which is a more realistic reflection of the continuous nature of ecological parameters.

2 Fuzzy Classification and Spatial Data Analysis

Classification and geostatistic belong to the main problems of ecological data analysis. Fuzzy clustering methods can be applied for fuzzy classification which means the partition of objects into classes with not sharply formed boundaries. The usual sharp cluster analysis, which definitely places an object within only one cluster, is not particularly useful for data of high uncertainty. With fuzzy clustering it is no longer essential to definitely place an object within one cluster, since the membership value of this object can be split up between different clusters. This enable us a better interpretation of data structure. We can find many applications of fuzzy clustering in different topics of ecology, e.g. Zang applies the fuzzy approach to the classification of ecological habitats [14] and Pedersen to examine the floristic and environmental similarity among reaches [12].

The common fuzzy clustering methods, like the fuzzy c-means method, work only with crisp data. In ecology we have often to deal with data with a semblance of

accuracy. In such cases the extension of the fuzzy c-means method for fuzzy data can be very helpful for the classification of imprecise ecological data. This extension was a basis for the development of the Fuzzy Clustering System EcoFucs v. 5.1 at the University of Kiel [9]. The cluster analysis of ecotoxicological data presented in this paper is used to illustrate the extension of the fuzzy c-means method for fuzzy data.

Spatial data is an essential part of ecological data. A fuzzy approach can be very useful for spatial data analysis when probabilistic approaches are inappropriate or impossible, e.g. for the classification of topo-climatic data [6] or for the land use mapping [10]. Fuzzy classification is now widely accepted in remote sensing of spatial data [11].

The fuzzy extension of the most popular interpolation procedure for spatial data, the so-called fuzzy kriging, is presented in this paper as an example of a fuzzy approach to spatial data analysis. The fuzzy kriging procedure can be used for the regionalization of ecological parameters. The implementation of the fuzzy kriging method is a part of the Fuzzy Kriging and Evaluation System FUZZEKS developed at the University of Kiel [3]. FUZZEKS utilizes exact (crisp) measurement data as well as imprecise estimates defined as fuzzy numbers and can be very useful for the analysis of spatial data in the case if an application of the conventional kriging method is restricted owing to insufficient amount of exact (crisp) data.

Fuzzy classification and fuzzy regionalization were used in this paper as an example of the suitability of a fuzzy approach to ecological data analysis.

3 Fuzzy Classification of Fuzzy Data

The fuzzy cluster analysis is a method of partitioning the n-dimensional space into clusters with not sharply formed boundaries. But the problem is, that the most common clustering methods, like the fuzzy c-means method by Bezdek [4], are only designed for treating crisp data, that means they provide the fuzzy partition only for crisp data (e.g. exact measurement data). Ecological data are often presented with a semblance of accuracy when exact values cannot be ascertained. Such problems naturally arise in applications when data are imprecise and information is not available about distributions of variances which describe data inaccuracy. In such cases it may only be possible to obtain estimates of data scatter which can be treated in the context of fuzzy sets and used for defining fuzzy data in the form of fuzzy vectors in a high dimension [7]. Yang defined the distance between two so-called conical fuzzy vectors $\widetilde{A}$ and $\widetilde{C}$ as follows [13]:

$$d_c^2\left(\widetilde{A},\widetilde{C}\right)=\left\|\overline{a}-\overline{c}\right\|^2+tr\left(\left(A-C\right)^T\left(A-C\right)\right) \qquad (1)$$

where:

A and C are the so-called panderance matrices of $\widetilde{A}$ and $\widetilde{C}$,

$\left\|\overline{a}-\overline{c}\right\|_{AC}$ is the distance norm (metric) between $\widetilde{A}$ and $\widetilde{C}$, and

the trace $tr\left(\left(A-C\right)^T\left(A-C\right)\right)$ is the diagonal sum of $\left(\left(A-C\right)^T\left(A-C\right)\right)$.

The panderance matrix of the conical fuzzy vector describes the accuracy of data. This matrix contains spreads of data on its diagonal.

Yang proofed that $d_c\left(\widetilde{A},\widetilde{C}\right)$ defined above is a complete metric, which is an assumption of the convergence of the fuzzy c-means clustering procedure by Bezdek. That means, we can define the well known objective function of the fuzzy c-means procedure for conical fuzzy vectors by

$$F\left(c\right)=\sum_{i=1}^{n}\sum_{j=1}^{c}\left(\mu_{ij}\right)^m d_c^2\left(\widetilde{A}_i,\widetilde{C}_j\right) \qquad (2)$$

where $\widetilde{A}_i$ is the *i*th object and $\widetilde{C}_j$ is the *j*th cluster, both defined as conical fuzzy vectors.

The clustering algorithm for conical vectors proposed by Yang has been extended for the diagonal norm using the so-called z-transformation of the Euclidian norm and implemented for the Fuzzy Clustering System EcoFucs v.5.1. The diagonal norm is a highly recommendable distance measure in the case of heterogeneous ecological data with different domain scales. In such cases we can transform data in a uniform manner before we start the fuzzy c-means procedure for conical vectors [9]:

$$\text{z-trans}(\widetilde{A}_i)=\frac{\widetilde{A}_i-\widetilde{M}}{\widetilde{S}} \qquad (3)$$

where:

$\widetilde{M}=\frac{1}{n}\sum_{i=1}^{n}\widetilde{A}_i$ is the mean vector of all fuzzy conical vectors $\widetilde{A}_i$ of the input data set, and

$\widetilde{S}$ is the vector of spreads from the panderance matrix.

To obtain back the coordinates of the cluster centers in the real scale we have to proceed the inverse transformation of the results of the fuzzy c-means procedure. EcoFucs offers four distance norms as a measure of similarity between the object and the respective clusters (the euclidean-, diagonal-, mahalonobis- and the L1-norm) for crisp data and the euclidean and diagonal norms for fuzzy vectors. This system offers also a set of methods for calculating the start partition (WARD, conventional c-means, maximum-distance-algorithm, sharp or fuzzy random partitions). The choice of the distance norm depends on the data set. The partition efficiency indicators available in EcoFucs (entropy, partition coefficient, payoff and

non-fuzziness index) can be very helpful in searching for the optimal partition.

3.1 An application example: A fuzzy cluster analysis of chemicals according to their ecotoxicological properties.

The fuzzy cluster analysis of chemicals according to their ecotoxicological properties has been used to check and illustrate the suitability of the fuzzy clustering procedure based on conical fuzzy vectors and to compare this extension with the fuzzy clustering procedure for crisp data. The example presented in this paper is particularly useful for this comparison because of the high uncertainty of ecotoxicological data. The uncertainty of these data arises from imprecision, a difficult comparability and a mixture of quantitative and qualitative data. In comparison to the conventional "sharp" clustering methods both fuzzy clustering techniques (for crisp and fuzzy data) are more appropriate to handle data of high uncertainty.

The data set for this example is derived from the reports of the Advisory Committee on Existing Chemicals of Environmental Relevance which selected and evaluated existing chemicals according to their environmental relevance [5]. Data shown in Table 3 (Appendix) is a part of the data set with about two hundreds chemicals used in [8] for the fuzzy clustering procedure for crisp data. These data can be divided into three groups of features relevant to ecotoxicological properties:

- Data relevant for the distribution of a compound between different compartments (log P_{OW} and Henry constant; columns 1 and 2 in the Table 3, Appendix),
- Data representing the potentials for biodegradability, hydrolysis and photolysis (columns 3 to 5), and
- Toxicity indicators: toxicity for a) microbes, b) aquatic invertebrates, c) aquatic vertebrates, d) mammals and a combined indicator term for cancerogeneity, mutageneity and teratogeneity (columns 6 to 10).

Both the costs of ecotoxicological testing procedures and the large number of existing chemicals make it necessary to select representative chemicals which faithfully reflect the relevant properties of possibly a major group of compounds. Therefore the main tasks of this analysis were to find distinguishable clusters with characteristic properties and to find chemicals representative for each cluster. The results of clustering of crisp data from Table 3 (Appendix) in 5 clusters for the diagonal norm are presented in Table 1. 5 clusters has been chosen as the "optimal" number of clusters by means of the analysis of the partition efficiency indicators offered by EcoFucs. The distribution of the membership values provides additional information from which the degree of similarity between properties of a particular chemical and properties characterising particular clusters can be deduced. This is particularly important since there is a lot of chemicals with more or less overlapping properties (see Table 1, the membership values with a membership ≥ 0.10 to different clusters are underlined) which would not be registered by conventional "sharp" clustering methods.

Table 1: The results of clustering of crisp data in 5 clusters for the diagonal norm (the numbers in bold-face show the highest membership values; membership values with a membership ≥ 0.10 to different clusters are underlined).

	Membership degrees to clusters:				
	1	2	3	4	5
Cluster 1					
Dichlorobenzene	**0,97**	0,03	0,00	0,00	0,00
Nonylphenol	**0,82**	0,16	0,01	0,00	0,00
124-Trichlorobenz.	**0,80**	0,16	0,02	0,00	0,00
Ditolylether	**0,76**	0,11	0,05	0,01	0,05
Dibutylphthalat	**0,64**	0,15	0,08	0,03	0,09
N-Ethylaniline	**0,90**	0,08	0,00	0,00	0,00
Chlorotoluidine	**0,56**	0,33	0,05	0,01	0,03
4-Nitrophenol	**0,71**	0,18	0,02	0,01	0,06
Cluster 2					
Chloroform	0,06	**0,91**	0,01	0,00	0,01
Pentachlorophenol	0,29	**0,67**	0,01	0,00	0,01
p-Nitromethoxybenz.	0,01	**0,98**	0,00	0,00	0,00
Benzene	0,18	**0,55**	0,10	0,04	0,11
o-Chlorotoluol	0,05	**0,92**	0,00	0,00	0,00
Nitrobenzene	0,04	**0,94**	0,00	0,00	0,00
1,2,4,5-Tetrachlorob.	0,30	**0,35**	0,16	0,07	0,10
Triethylentetramine	0,20	**0,30**	0,18	0,07	0,24
Cluster 3					
Diphenylamine	0,20	0,16	**0,53**	0,06	0,03
o-Tolidine	0,00	0,00	**0,98**	0,00	0,00
o-Dianisidine	0,00	0,00	**0,98**	0,00	0,00
3,3Dichlorobenzidine	0,18	0,11	**0,60**	0,03	0,06
Cluster 4					
Trichloromethylben.	0,01	0,01	0,01	**0,95**	0,01
Benzoylchloride	0,03	0,03	0,06	**0,80**	0,06
Cluster5					
Diethylenglykoldim.	0,06	0,06	0,21	0,03	**0,61**
Hexanedioicacid	0,02	0,02	0,01	0,01	**0,92**
Aceticacidanhydride	0,05	0,05	0,02	0,03	**0,83**
N,N-Dimethylform.	0,04	0,05	0,13	0,03	**0,72**

The classification of chemicals presented in Table 1 is very similar to the results in [8]. Four clusters (1, 3, 4 and 5) have well distinguishable properties. It can be recognized by the location of the cluster centers shown in Table 4 (Appendix), e.g. cluster 4 is characterized by extremely high hydrolysis rates, high photolysis rates and high biodegradability. The one cluster, namely the cluster 2, contains the "rest" of chemicals with no significant common properties. They cannot be

definitely classified to one of other clusters with a clear characteristic.

The next step of this cluster analysis was the "fuzzification" of crisp data from the Table 3, that means they are defined now as conical fuzzy vectors (an example in Table 5, Appendix). The first line of each chemical in Table 5 contains the same crisp data, but now we interpret them as apexes of conical fuzzy vectors. The second line are spreads which illustrate the data accuracy for each features.

Table 2. The results of clustering of conical fuzzy vectors in 5 clusters for the diagonal norm; the names in bold-face show chemicals which changed their location from the "rest"-cluster 2 to other clusters.

	Membership degrees to clusters:				
	1	2	3	4	5
Cluster 1					
Dichlorobenzene	**0,88**	0,06	0,03	0,00	0,03
Nonylphenol	**0,67**	0,28	0,02	0,01	0,02
124-Trichlorobenzene	**0,55**	0,35	0,05	0,02	0,03
Ditolylether	**0,81**	0,08	0,03	0,01	0,05
Dibutylphthalat	**0,75**	0,08	0,04	0,03	0,08
N-Ethylaniline	**0,95**	0,03	0,00	0,00	0,01
Chlorotoluidine	**0,79**	0,09	0,02	0,03	0,06
4-Nitrophenol	**0,58**	0,16	0,06	0,04	0,15
Cluster 2					
Pentachlorophenol	0,08	**0,88**	0,02	0,01	0,01
p-Nitromethoxyben.	0,08	**0,88**	0,01	0,01	0,02
o-Chlorotoluol	0,02	**0,85**	0,01	0,09	0,02
Nitrobenzene	0,02	**0,96**	0,01	0,00	0,01
1,2,4,5-Tetrachlorob.	0,23	**0,30**	0,17	0,14	0,15
Cluster 3					
Diphenylamine	0,32	0,18	**0,32**	0,09	0,09
Benzene	0,17	0,22	**0,22**	0,20	0,19
o-Tolidine	0,00	0,00	**0,99**	0,00	0,01
o-Dianisidine	0,00	0,00	**0,99**	0,00	0,01
3,3Dichlorobenzidine	0,20	0,09	**0,55**	0,06	0,10
Cluster 4					
Trichloromethylbenz.	0,01	0,00	0,01	**0,97**	0,01
Benzoylchloride	0,12	0,06	0,14	**0,37**	0,31
Cluster5					
Chloroform	0,30	0,24	0,10	0,06	**0,30**
Diethylenglykoldim.	0,07	0,03	0,14	0.03	**0,73**
Hexanedioicacid	0,04	0,02	0,03	0,02	**0,90**
Aceticacidanhydride	0,10	0,04	0,03	0,06	**0,77**
N,N-Dimethylform.	0,03	0,01	0,06	0,02	**0,87**
Triethylentetramine	0,21	0,16	0,19	0,08	**0,36**

The results of the clustering procedure for conical fuzzy vectors show a very similar classification of chemicals to the clusters 1, 3, 4 and 5, but the number of compounds of the "rest"-cluster 2 is significantly reduced (Table 2). These chemicals could not be clearly classified by clustering of crisp data. Chloroform and Triethylentetramine changed their location to cluster 5 and Benzene to cluster 3. They all still have a high membership degree to cluster 2, but now their characteristic properties can be better recognized.

4 Fuzzy Regionalization

Fuzzy kriging is an extension of the conventional kriging procedure which is usually used for regionalization of spatial data [2]. Kriging is a common interpolation method which is based on a statistical analysis of spatial data. The first step of this analysis is the preparation of the so-called experimental variogram following by the fitting of the theoretical variogram, which is a basis for the interpolation procedure. The application of the conventional methods of spatial interpolation is often restricted owing to an insufficient amount of data. If the collection of new data is too expensive or impossible, we can consider the use of additional imprecise data subjectively estimated by an expert. We can use fuzzy sets to handle the imprecision and uncertainty of these data. In such a case the interpolated value will be a fuzzy number if at least one of input values is a fuzzy value defined by an expert.

The main kriging estimation (4) is a linear combination of the input values and can be calculated using the extension principle of the fuzzy set theory and the α-cut-representation of fuzzy sets [1]:

$$Z^*(x)_\alpha = \sum_{i=1}^{n} \delta_i(x) Z(x_i)_\alpha \qquad (4)$$

where:

$Z^*(x)_\alpha$ is the α-cut of the interpolated value $Z^*(x)$ at the position x,

$Z(x_i)_\alpha$ are the α-cuts of the input values $Z(x_i)$, and

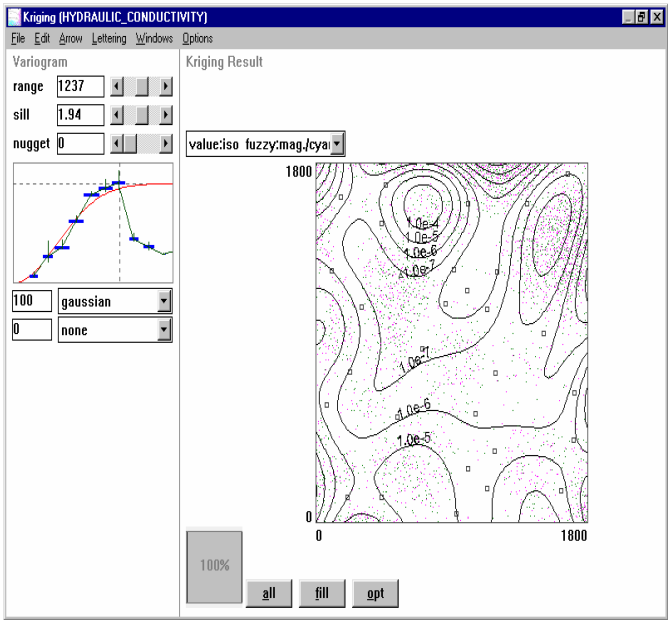
$\delta_i(x)$ are the crisp minimizing parameters.

The kriging estimation formulated by α-cut-representation of fuzzy sets (the equation 4) has been used for the implementation of the fuzzy kriging procedure for the Fuzzy Kriging and Evaluation System FUZZEKS developed at the University of Kiel. The implemented fuzzy kriging procedure utilizes exact (crisp) measurement data as well as imprecise estimates obtained from an expert. These imprecise data can be defined as fuzzy numbers and taken as additional input data for the kriging procedure.

Fig. 1 presents the main kriging window of FUZZEKS. The system supports a user in an interactive fitting of the crisp theoretical variogram to the fuzzy experimental variogram (see the small window on the left in Fig. 1). The results can be viewed as a map (isolines of interpolated values, see the window on the right in Fig. 1) or as a fuzzy number at any particular point of the area under investigation, or as a distribution of the membership function along a section line between two given points. Fig. 1 is an illustration which presents the results of fuzzy kriging of hydraulic conductivity for the data set with 26 crisp measurement points and 3 estimated values given in the area under investigation. The density of small points within the window on the right illustrates the degree of fuzziness of the results.

To simplify the preparation of the input data file a special ASCII - file format was implemented, combining both exact (crisp) and fuzzy data (fuzzy numbers) in one unified form. The input data can be also transformed (e.g. by a logarithmic function) before the calculation of the experimental variogram.

Fig. 1: The kriging window of FUZZEKS. Right window: isolines of interpolated values; the density of small points illustrates the degree of fuzziness of the results. Left window: interactive fitting of the theoretical variogram to the experimental variogram.



5 Final Remarks / Conclusions

The goal of this paper was to illustrate the suitability of a fuzzy approach to some problems of ecological data analysis. Fuzzy representation and interpretation of data structure is a very natural and intuitively plausible way to formulate and to solve some uncertainty problems in environmental data analysis.

A fuzzy partition of ecological objects obtained by fuzzy clustering methods presented in this paper reflects very well the continuous nature of ecological features. The clustering of data in the form of conical fuzzy vectors does not force one to describe objects with a semblance of accuracy and can be very helpful for the analysis of data of high imprecision. The conical fuzzy vectors are better suitable to illustrate the overlapping properties of classified objects.

The analysis of spatial data belongs to the main problems of the analysis of ecological data. In the case of an insufficient amount of data the extension of the data set by additional data subjectively estimated by an expert is particularly important. The fuzzy kriging procedure utilizes crisp measurement data as well as fuzzy data and can be used as a regionalization tool. The development of easy-to-use tools (like FUZZEKS or EcoFucs) for research or for practical tasks is very important for the promotion of fuzzy logic applications in ecology.

References:

[1] Bandemer, H., Gottwald, S., *Einführung in Fuzzy-Methoden*. Akademie Verlag, 1993.

[2] Bárdossy A., Bogardi I., and Kelly W. E., Geostatistics utilizing imprecise (fuzzy) information., *Fuzzy Sets and Systems*, 31/3, 1989, pp. 311-327

[3] Bartels F., *Ein Fuzzy-Auswertungs- und Krigingsystem für raumbezogene Daten*. Diplomarbeit, Inst. für Informatik und Praktische Mathematik, Universität Kiel, 1997.

[4] Bezdeck, J.C., A convergence theorem for the fuzzy c-means clustering algorithm, *IEEE Trans. PAMI*, PAMI2(1), 1980, pp. 1-8.

[5] BUA Beratungsgremium für umweltrelevante Altstoffe der GDCh (Editors), *BUA-Stoffberichte 1-90,* 1993.

[6] Burrough, P.A., Wilson, J.P., van Gaans, P.F.M. and Hansen, A.J., Fuzzy k-means classification of topo-climatic data as an aid to forest mapping in the Greater Yellowstone Area, USA, *Landscape Ecology*, 16/6, 2001, pp. 523-546.

[7] Celmiņš, A., Least Squares model fitting to fuzzy vector data, *Fuzzy Sets an Systems*, 22, 1987, pp. 245-269.

[8] Friderichs, M., Fränzle, O., Salski, A., Fuzzy clustering of existing chemicals according to their ecotoxicological properties, *Ecological Modelling*, 85, 1996, pp. 27-40.

[9] Gömann, M., *Erweiterung des Fuzzy-Clustering-Systems EcoFucs*, Studienarbeit, Institut für Informatik, Christian-Albrechts-Universität zu Kiel, 2002.

[10] Liu, M. and Samal, A. 2002. A fuzzy clustering approach to delineate agroecozones. *Ecological Modelling*, 149/3, 2002, pp. 215-228.

[11] Malins, D. and Metternicht, G., 2006. Assessing the spatial extent of dryland salinity through fuzzy modelling, *Ecological Modelling*, 193/3-4, 2006, pp. 387-411.

[12] Pedersen T.C.M., Baattrup-Pedersen, A. and Madsen, T.V., 2006. Effects of stream restoration and management on plant communities in lowland streams, *Freshwater Biology*, 51, 2006, pp. 161-179.

[13] Yang, M.-S. and Liu, H-H, Fuzzy clustering procedures for conical fuzzy vector data, *Fuzzy Sets and Systems*, 106, 1999, pp. 189-200.

[14] Zang, L., Liu, Ch., Davis, C.J.,Solomon, D.S., Brann, T.B. and Caldwell, L.E., Fuzzy classification of ecological habitats from FIA data, *Forest Science*, 50/1, 2004, pp. 117-127.

Appendix

Table 3. The crisp data set used as testing data for the fuzzy cluster analysis of chemicals according to their ecotoxicological properties.

Features:1: log P_{OW} *6: toxicity for microbes* (score)

2: Henry constant [Pa m³ mol^{-1}] *7: toxicity for aquatic invertebrates* (score)

3: biodegradability (score) *8: toxicity for aquatic vertebrates* (score)

4: degradation in water [1/ t½] in days *9:mammalian toxicity* (score)

5: degradation in the air [1/ t½] in days 1*0:cancerogeneity/mutageneity/teratogeneity (score)*

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Chlorotoluidine	2.58	0.1	7	0	8	10	0	3	10	5
Triethylentetramine ----------	-1.44	0.0	0	0	14.04	20	3	0	10	5
Chloroform----------	1.9	310	3	0	0.0083	10	3	10	3	5
Pentachlorophenol----------	3.0	0.04	2	1	1	0	0	0	0	10
Dichlorobenzene----------	3.60	3.1	2	0	0.04	3	3	3	10	10
p-Nitromethoxybenzene-------	2.03	0.2	2	0	0.196	10	3	3	0	5
Nonylphenol----------	3.27	136	2	0	0	3	0	0	10	5
Diphenylamine----------	3.62	0.09	2	0	24	3	3	10	10	5
124-Trichlorobenzene---------	4.21	1.08	2	0	0.033	0	3	0	10	5
Ditolylether----------	5.54	15.7	2	0	0	10	3	3	20	15
Dibutylphthalat----------	4.7	0.27	10	0.0003	0.56	10	3	0	20	5
Benzene----------	2.1	440	10	0	0.077	10	10	3	0	0
o-Tolidine----------	2.4	0.00	3	0.33	12	3	20	20	10	0
o-Dianisidine----------	2.0	0.00	2	0	12	3	20	20	10	0
3,3Dichlorobenzidine----------	3.55	0.00	0	0	12	10	20	0	20	0
o-Chlorotoluol----------	3.45	970	0	0	0.18	10	3	3	0	5
N-Ethylaniline----------	3.0	5.0	3	0	0.003	10	3	3	10	10
Nitrobenzene----------	1.6	2.4	3	0.04	0.014	3	3	3	0	5
Diethylenglykoldimethyleth.	-0.36	0.04	2	0	1	20	20	20	20	5
Hexanedioicacid----------	0.08	0.10	10	0	0.227	20	20	10	10	20
Aceticacidanhydride----------	-0.2	0.0	10	232	0.045	20	3	10	10	15
Trichloromethylbenzene-----	2.92	39.2	10	655	20	10	3	20	3	0
Benzoylchloride----------	0.0	0.0	10	400	20	10	20	3	10	5
4-Nitrophenol----------	1.9	0.00	7	0	0.143	3	3	3	10	15
N,N-Dimethylformamide-----	-1.01	0.00	10	0	0.25	10	20	20	10	5
1,2,4,5-Tetrachlorobenzene--	4.6	15290	2	0.02	0.0065	3	3	3	3	5

Table 4. The coordinates of the cluster centers obtained by clustering of crisp data from the Table 3 in 5 clusters.

cluster.1:	03,571	551,764	03,618	07,945	01,636	05,754	02,628	01,999	10,864	08,829
cluster 2:	02,499	885,509	02,853	07,670	01,499	07,395	03,168	03,600	02,230	05,793
cluster.3:	02,436	332,654	02,304	03,103	12,417	05,283	16,664	14,377	11,571	01,306
cluster 4:	01,651	216,638	09,607	508,186	19,164	10,029	10,304	12,011	06,306	02,448
cluster 5:	-00,151	169,385	08,101	60,164	01,122	17,224	14,347	12,785	11,542	11,866

Table 5. An example of conical fuzzy vectors defined for the first two chemicals from the crisp data from the Table 3. The first line of each chemical contains the apexes and the second line determinates the spreads of each feature.

Chlorotoluidine ----------	2.58	0.1	7	0	8	10	0	3	10	5
	0.01	0.01	2	0.01	0.01	5	0.01	2	5	5
Triethylentetramine------	-1.44	0	0	0	14.04	20	3	10	10	5
	0.01	0.01	0.01	0.01	0.01	5	2	5	5	5