

PROBABILISTIC AND STATISTICAL FUZZY SET FOUNDATIONS OF
COMPETITIVE EXCEPTION LEARNING

J. VAN DEN BERG, W.M. VAN DEN BERGH, U. KAYMAK

ERIM REPORT SERIES <i>RESEARCH IN MANAGEMENT</i>	
ERIM Report Series reference number	ERS-2001-40-LIS
Publication	July 2001
Number of pages	10
Email address first author	jvandenber@few.eur.nl
Address	Erasmus Research Institute of Management (ERIM) Rotterdam School of Management / Faculteit Bedrijfskunde Erasmus Universiteit Rotterdam P.O. Box 1738 3000 DR Rotterdam, The Netherlands Phone: +31 10 408 1182 Fax: +31 10 408 9640 Email: info@erim.eur.nl Internet: www.erim.eur.nl

Bibliographic data and classifications of all the ERIM reports are also available on the ERIM website:
www.erim.eur.nl

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT

REPORT SERIES *RESEARCH IN MANAGEMENT*

BIBLIOGRAPHIC DATA AND CLASSIFICATIONS		
Abstract	<p>Recently, a Competitive Exception Learning Algorithm (CELA) was introduced [1, 2]. This algorithm establishes an optimal mapping from a (continuous) M-dimensional input sample space to an N-dimensional (continuous) output sample space. CELA is aimed to discover regimes (i.e. local behavior in the input sample space) for which the conditional probability distribution in the output sample space systematically deviates from the average unconditional distribution. Previous papers on CELA dealt with the introduction of the algorithm by sketching its background and by describing the algorithmic sub-steps. The algorithm was tested successfully on both simulated and real world data, mainly in the field of financial markets. However, until now a precise and firm theoretical foundation of CELA is still lacking. The current paper resolves this imperfection. The contribution to be made here is twofold. First, we present, in section 2, a probability theory and statistics of fuzzy sets which in itself is interesting. Second, we re-formulate, in section 3, the CELA-algorithm within the probabilistic fuzzy framework introduced. We finalize with a discussion and outlook.</p>	
Library of Congress Classification (LCC)	5001-6182	Business
	5201-5982	Business Science
	QA 273-274.76	Probabilities
Journal of Economic Literature (JEL)	M	Business Administration and Business Economics
	M 11	Production Management
	R 4	Transportation Systems
	C 19	Econometric and Statistical Methods:Other
European Business Schools Library Group (EBSLG)	85 A	Business General
	260 K	Logistics
	240 B	Information Systems Management
	250 C	Statistics
Gemeenschappelijke Onderwerpsontsluiting (GOO)		
Classification GOO	85.00	Bedrijfskunde, Organiseatiekunde: algemeen
	85.34	Logistiek management
	85.20	Bestuurlijke informatie, informatieverzorging
	31.70	Waarschijnlijkheidsrekening
Keywords GOO	Bedrijfskunde / Bedrijfseconomie	
	Bedrijfsprocessen, logistiek, management informatiesystemen	
	Waarschijnlijkheidsleer, Fuzzy Set, Algoritmen	
Free keywords	Probability Theory, Statistics, Fuzzy Set, Exception Learning	

Probabilistic and Statistical Fuzzy Set Foundations of Competitive Exception Learning

J. van den Berg W.M. van den Bergh
U. Kaymak

Erasmus University Rotterdam, Faculty of Economics

Dept. of Computer Science, room H9-24, P.O. Box 1738, 3000 DR, Rotterdam, the Netherlands

tel: +31-10-4081342, +31-10-4081286, +31-10-4081350, fax: +31-10-4089167

e-mail: jvandenbergh@few.eur.nl, vandenbergh@few.eur.nl, u.kaymak@ieee.org

Abstract

Keywords: Probability Theory, Statistics, Fuzzy Set, Exception Learning.

Technical area of the paper: T5 — Mathematics.

1 Introduction

Recently, a Competitive Exception Learning Algorithm (CELA) was introduced [1, 2]. This algorithm establishes an optimal mapping from a (continuous) M -dimensional input sample space to an N -dimensional (continuous) output sample space. CELA is aimed to discover regimes (i.e. local behavior in the input sample space) for which the conditional probability distribution in the output sample space systematically deviates from the average unconditional distribution. Previous papers on CELA dealt with the introduction of the algorithm by sketching its background and by describing the algorithmic sub-steps. The algorithm was tested successfully on both simulated and real world data, mainly in the field of financial markets. However, until now a precise and firm theoretical foundation of CELA is still lacking. The current paper resolves this imperfection. The contribution to be made here is twofold. First, we present, in section 2, a probability theory and statistics of fuzzy sets which in itself is interesting. Second, we re-formulate, in section 3, the CELA-algorithm within the probabilistic fuzzy framework introduced. We finalize with a discussion and outlook.

2 Probability Theory and Statistics on Fuzzy Sets

In this section, we introduce the theory needed to be able to explain the background of the CELA in the next session. The theory introduced here is strongly related to the approach presented in the seminal

paper [3]. We made further use of several standard works, namely on probability theory [4], mathematical statistics [5], and fuzzy set theory [6]. We only explain the theory for one-dimensional spaces. However, this can easily be extended to multi-dimensional spaces.

2.1 The discrete case

2.1.1 Classical probability theory

Let us start considering classical probability theory defined on crisp sets. More specifically, we consider a random variable \underline{x} defined on a discrete sample space $X = \{x_1, x_2, \dots\}$ with sample points x_1, x_2, \dots . Then the probability of sample point x_i is often written as

$$\Pr(\underline{x} = x_i) = f(x_i) \quad (1)$$

with various properties, for example, $\sum_i f(x_i) = 1$. A set $A \subseteq X$ of sample points x_i , denoted as $A = \{x_i, x_j, \dots\}$, is often called an *event* [4]. The probability $\Pr(A)$ of such an event is defined as the sum of the probabilities of all sample points in it, so

$$\Pr(A) = \sum_{x_i \in A} f(x_i). \quad (2)$$

We also shall make use of the characteristic function $\chi_B(x)$ which defines a set B as follows:

$$\chi_B(x) = \begin{cases} 1 & \text{if } x \in B \\ 0 & \text{if } x \notin B. \end{cases} \quad (3)$$

We next consider the special events S_i consisting of just one single sample point, so $\forall i : S_i = \{x_i\}$. We call these events *singleton events*. From (3) we conclude that

$$\chi_{S_i}(x) = \begin{cases} 1 & \text{if } x = x_i \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

By applying equations (1), (2) and (4), we conclude that

$$\Pr(S_i) = \Pr(\underline{x} = x_i) = f(x_i) = \chi_{S_i}(x_i)f(x_i) = \sum_{x_k \in X} \chi_{S_i}(x_k)f(x_k). \quad (5)$$

Note that $\sum_{x_k \in X}$ means that we sum over *all* sample points x_k in X . Similarly and more generally, we derive that the probability of an arbitrary event $A = \{x_i, x_j, \dots\}$ in our sample space can be written as

$$\Pr(A) = \sum_{x_i \in A} f(x_i) = \sum_{x_k \in X} \chi_A(x_k)f(x_k) = E(\chi_A(\underline{x})) \quad (6)$$

where $E(\underline{x})$ is the classical mean or mathematical expectation of \underline{x} . It should further be clear that the characteristic function $\chi_A(\underline{x})$ is a stochastic variable: this is a direct consequence of the fact that \underline{x} is stochastic.

2.1.2 Probability theory on fuzzy sets

The above-given theory can easily be generalized to a probability theory on fuzzy sets (also called a probabilistic fuzzy set theory). We first introduce the notion of a *fuzzy singleton event*: like a single

(crisp) event $S_i = \{x_i\}$ of X can be characterized by its corresponding characteristic function $\chi_{S_i}(x)$ as defined by (4), so a fuzzy singleton event can be described by a membership function $\mu_{S_i}(\underline{x})$

$$\mu_{S_i}(x) = \begin{cases} m_{S_i}(x_i) & \text{if } x = x_i \\ 0 & \text{if } x = x_j \text{ with } j \neq i, \end{cases} \quad (7)$$

where $0 \leq m_{S_i}(x_i) \leq 1$. In addition and more generally, a discrete *fuzzy event* A can be characterized by the membership function $\mu_A(x)$:

$$\forall x_k \in X : \mu_A(x_k) = m_A(x_k), \quad (8)$$

with $\forall x_k : 0 \leq m_A(x_k) \leq 1$.

The probability of a discrete fuzzy singleton event can be defined by generalizing expression (5): we simply replace the characteristic function $\chi_{S_i}()$ by $\mu_{S_i}()$ and get

$$\Pr(S_i) = \mu_{S_i}(x_i)f(x_i) = \sum_{x_k \in X} \mu_{S_i}(x_k)f(x_k). \quad (9)$$

More generally, the probability of a discrete fuzzy event A is defined by

$$\Pr(A) = \sum_{x_k \in X} \mu_A(x_k)f(x_k) = E(\mu_A(\underline{x})). \quad (10)$$

Hence, the probability of a fuzzy event equals the mathematical expectation of the membership function describing this fuzzy event. Note that here the membership function $\mu_A(\underline{x})$ is stochastic.

2.1.3 Well-defined fuzzy sample spaces

Next, we consider the case that various fuzzy events A_c ($c = 1, \dots, C$) are defined on our discrete sample space conform definition (8), so

$$\forall A_c \forall x_k \in X : \mu_{A_c}(x_k) = m_{A_c}(x_k). \quad (11)$$

So, every sample point x_k corresponds to a vector of membership values $(\mu_{A_1}(x_k), \mu_{A_2}(x_k), \dots, \mu_{A_C}(x_k))$. Within this setting, we interpret each x_k as a ‘fuzzy sample point’ or, shortly, a *fuzzy sample*. To explain, in a crisp sample space, any sample x_k does belong to an event A or it does not belong to that event. In a fuzzy sample space however, each x_k belongs to several different fuzzy events to a certain degree. As will be explained below, we are interested in a very special sample space X of fuzzy events, namely one for which the membership values of each fuzzy sample point sum up to one, so that $A_c, c = 1, \dots, C$ form a fuzzy partition [7] of X . In mathematical terms, we suppose that

$$\forall x_k : \sum_{A_c} \mu_{A_c}(x_k) = 1. \quad (12)$$

If the last equation holds, we can proof an interesting property of a discrete fuzzy sample space.

Theorem 2.1 *Let a set of fuzzy events $A_1, A_2, \dots, A_c, \dots$ be given where each event A_c is described by its membership function $\mu_{A_c}(x)$. Let in addition, the membership values of each fuzzy sample x_k comply with condition (12), then the sum of all probabilities of fuzzy events A_c equals one, or, in mathematical terms*

$$\sum_{A_c} \Pr(A_c) = 1 \quad (13)$$

Proof Using definition (10) and condition (12), we conclude that

$$\sum_{A_c} \Pr(A_c) = \sum_{A_c} \sum_{x_k \in X} \mu_{A_c}(x_k) f(x_k) = \sum_{x_k \in X} f(x_k) \sum_{A_c} \mu_{A_c}(x_k) = \sum_{x_k \in X} f(x_k) = 1. \quad (14)$$

□

Theorem (2.1) shows that condition (12) guarantees that the probabilities on the fuzzy events in the fuzzy sample space, sum up to precisely one. In other words, these conditions guarantee what we shall term a *well-defined* sample space.

2.2 The continuous case

2.2.1 Classical probability theory

Again we start considering classical probability theory defined on crisp sets. More specifically, we consider a random variable \underline{x} defined on a continuous sample space X . Within X , an interval $I_i = [x_{i-1}, x_i]$ defines an event and the probability of event I_i can be described as

$$\Pr(I_i) = \Pr(x_{i-1} \leq \underline{x} \leq x_i) = \int_{x_{i-1}}^{x_i} f(x) dx = \int_{x_{i-1}}^{x_i} \chi_{I_i}(x) f(x) dx = \int_{-\infty}^{\infty} \chi_{I_i}(x) f(x) dx, \quad (15)$$

where $f(x)$ is a probability density function. Next, we consider the more general event A consisting of the union of several disjunct (crisp) intervals I_i, I_j, \dots , so

$$A = I_i \cup I_j \cup \dots, \quad \text{where } \forall i \neq j : I_i \cap I_j = \emptyset. \quad (16)$$

Using this definition it is not very difficult to proof that

$$\chi_A(x) = \sum_{I_i \subseteq A} \chi_{I_i}(x). \quad (17)$$

Using (15) and (17), the probability of the crisp event A can be expressed as

$$\Pr(A) = \Pr(I_i \cup I_j \cup \dots) = \sum_{I_i \subseteq A} \Pr(I_i) = \sum_i \Pr(x_{i-1} \leq \underline{x} \leq x_i) = \quad (18)$$

$$\sum_{I_i \subseteq A} \int_{-\infty}^{\infty} \chi_{I_i}(x) f(x) dx = \int_{-\infty}^{\infty} \left(\sum_{I_i \subseteq A} \chi_{I_i}(x) \right) f(x) dx = \int_{-\infty}^{\infty} \chi_A(x) f(x) dx = E(\chi_A(\underline{x})). \quad (19)$$

2.2.2 Probability theory on fuzzy sets

Like in the discrete case, the theory of classical probability theory as explained in subsection 2.2.1 can easily be generalized. For an arbitrary *fuzzy interval* I_i defined on X by its membership function $\mu_{I_i}(x)$, we generalize (15) by replacing $\chi_A()$ by $\mu_A()$ and get

$$\Pr(I_i) = \int_{x_{i-1}}^{x_i} \mu_{I_i}(x) f(x) dx = \int_{-\infty}^{\infty} \mu_{I_i}(x) f(x) dx. \quad (20)$$

Similarly, for a *fuzzy event* A , defined on X by its membership $\mu_A(x)$, we generalize (18) and get

$$\Pr(A) = \int_{-\infty}^{\infty} \mu_A(x) f(x) dx = E(\mu_A(\underline{x})). \quad (21)$$

Like in the discrete case, the probability of a fuzzy event equals the mathematical expectation of the stochastic membership function describing this fuzzy event.

2.2.3 Well-defined fuzzy sample spaces

Theorems (2.1) also holds in the continuous case:

Theorem 2.2 *Let a set of fuzzy events $A_1, A_2, \dots, A_c, \dots$ be given where each event A_c is described by its membership function $\mu_{A_c}(x)$. Let in addition, the membership values of each fuzzy sample x comply with condition (12), then the sum of all probabilities of fuzzy events A_c equals one, or, in mathematical terms*

$$\sum_{A_c} \Pr(A_c) = 1 \quad (22)$$

Proof Using the above-given probability definitions and condition (12), we find

$$\sum_{A_c} \Pr(A_c) = \sum_{A_c} \int_{-\infty}^{\infty} \mu_{A_c}(x) f(x) dx = \int_{-\infty}^{\infty} f(x) \sum_{A_c} \mu_{A_c}(x) dx = \int_{-\infty}^{\infty} f(x) dx = 1. \quad (23)$$

□

Again condition (12) guarantees that the corresponding sample space is *well-defined* in the sense that the sum of the probabilities of the fuzzy events A_c equals precisely 1.

2.3 The statistical part

Having a set of P representative sample data x_p , $p = 1, \dots, P$, from a continuous sample space with a probability density function $f(x)$, then discrete mathematical statistics can be used to assess probabilities, both in the crisp and in the fuzzy case.

2.3.1 Assessing unconditional probabilities

In the crisp case of assessing probabilities, the domain of X is partitioned in a final set of (crisp) *classes* $A_c = [x_{c-1}, x_c]$. Let \tilde{f}_{A_c} denote the relative frequency and f_{A_c} the absolute frequency of sample points $x_p \in A_c$, then the probability of fuzzy class A_c can be assessed conform

$$\Pr(A_c) \approx \tilde{f}_{A_c} = \frac{f_{A_c}}{P} = \frac{1}{P} \sum_{x_p} \chi_{A_c}(x_p) = \hat{\chi}_{A_c}, \quad (24)$$

where $\hat{\chi}_{A_c}$ represents the average characteristic function. The vector $(\tilde{f}_{A_1}, \tilde{f}_{A_2}, \dots, \tilde{f}_{A_c})$ describes an assessment of the probability distribution over all classes A_c and forms in this way an empirical characterization of the sample space X .

In the fuzzy case, we can easily generalize the approach. This time, the domain of X is fuzzily partitioned in a final set of *fuzzy classes* A_c . This is here done in such a way that condition (12) concerning the membership functions $\mu_{A_c}(x)$, holds. This guarantees that X is a well-defined sample space. Let in this case \tilde{f}_{A_c} denote the relative frequency and f_{A_c} the absolute frequency of the contributions of the fuzzy samples x_p to the fuzzy class A_c (remember that each fuzzy sample belongs to each fuzzy class to a certain degree), then the probability of fuzzy class A_c can be assessed conform

$$\Pr(A_c) \approx \tilde{f}_{A_c} = \frac{f_{A_c}}{P} = \frac{1}{P} \sum_{x_p} \mu_{A_c}(x_p) = \hat{\mu}_{A_c}, \quad (25)$$

Comparing (21) and (25), we see that the mathematical mean $E(\mu_{A_c}(\underline{x}))$ of a fuzzy class or event A_c can be assessed by the empirical mean $\hat{\mu}_{A_c}$. Finally we observe that the vector $(\tilde{f}_{A_1}, \tilde{f}_{A_2}, \dots, \tilde{f}_{A_c})$ describes an assessment of the probability distribution over all fuzzy classes A_c and forms in this way an empirical characterization of the fuzzy sample space X .

2.3.2 Assessing conditional probabilities

In the crisp case, conditional probabilities $\Pr(A_c|A_b)$ can be assessed conform

$$\Pr(A_c|A_b) = \frac{\Pr(A_c \cap A_b)}{\Pr(A_b)} \approx \frac{\tilde{f}_{A_c \cap A_b}}{\tilde{f}_{A_b}} = \frac{\sum_{x_p} \chi_{A_c \cap A_b}(x_p)}{\sum_{x_p} \chi_{A_b}(x_p)} = \frac{\sum_{x_p} \chi_{A_b}(x_p) \chi_{A_c}(x_p)}{\sum_{x_p} \chi_{A_b}(x_p)} \quad (26)$$

where we use the property that $\chi_{A_c \cap A_b}(x_p) = \chi_{A_b}(x_p) \chi_{A_c}(x_p)$. Interpreting the frequencies used in the formulas as fuzzy frequencies, we can generalize (26) to

$$\Pr(A_c|A_b) = \frac{\Pr(A_c \cap A_b)}{\Pr(A_b)} \approx \frac{\tilde{f}_{A_c \cap A_b}}{\tilde{f}_{A_b}} = \frac{\sum_{x_p} \mu_{A_c \cap A_b}(x_p)}{\sum_{x_p} \mu_{A_b}(x_p)} = \frac{\sum_{x_p} \mu_{A_b}(x_p) \mu_{A_c}(x_p)}{\sum_{x_p} \mu_{A_b}(x_p)}. \quad (27)$$

3 Competitive Exception Learning

In this section, we re-formulate the Competitive Exception Learning Algorithm (CELA) within the probabilistic and statistical fuzzy framework introduced in the previous section. The extra assumption needed is that this framework also holds in multi-dimensional spaces.

3.1 Goal of CELA

The CELA tries to learn a mapping from a (continuous) M -dimensional input sample space X to an N -dimensional (continuous) output sample space Y . To do so, a representative set of fuzzy samples $(x_p, y_p), (p = 1, 2, \dots, P)$ is available generated by a probabilistic fuzzy data source, where $x_p \in X$ and $y_p \in Y$. It is assumed that the fuzzy samples y_p in Y are usually generated conform a certain probability distribution independent on the generation of the corresponding samples x_p in X . However, under certain special conditions called ‘regimes’, related to special values x_p observed in X , the fuzzy sample values y_p are generated conform a different probability distribution. In other words, the process of generating output fuzzy samples is assumed to be (more or less) *conditional* on the process of generating input fuzzy samples. *It is the main task of CELA to learn the regimes (i.e., the values x in X) for which the conditional probability distribution in Y deviates exceptionally (i.e., at most) from the average (i.e., the unconditional) probability distribution in Y .*

3.2 Fuzzy classes

In order to achieve the described goal of CELA, a set of B classes A_b within X , and a set of C classes A_c in Y is defined. Each class A_b and each class A_c has a parameter (called the class center) which should be assessed during execution of CELA. After assessing this parameter (for class A_b denoted as \bar{x}_b , for class A_c denoted as \bar{y}_c), the corresponding class is fixed. The idea behind the way of fixing of the output

classes A_c is to facilitate a compact, close assessment of the unconditional probability distribution in Y (details will follow below). The fixation of the input classes A_b serves a completely different goal: after fixation, an input class A_b is assumed to coincide with a regime (the above-introduced notion) after which the exceptional conditional probability distribution can be assessed quite easily (again, details will follow below).

The membership functions μ_{A_b} (μ_{A_c}) describing the fuzzy classes A_b (A_c) should meet several requirements. First, they should be defined locally, i.e., class A_b (A_c) should have high membership values for x -values (y -values) near class center \bar{x}_b (\bar{y}_c) and low membership values for x -values (y -values) near other class centers. Second, the membership functions should meet condition (12) in order to guarantee that the fuzzy sample space X (Y) is ‘well-defined’ (see sections 2.1.3 and 2.2.3). These requirements motivated us to define the membership functions as follows:

$$\mu_{A_b}(x) = \frac{d_b^{-q}(x)}{\sum_{k=1}^B d_k^{-q}(x)} \quad \text{and} \quad \mu_{A_c}(y) = \frac{d_c^{-q}(y)}{\sum_{k=1}^C d_k^{-q}(y)}. \quad (28)$$

The letter ‘d’ in these definitions represent a distance measure where $d_b(x)$ represents the (e.g., Euclidean) distance between x and center \bar{x}_b in X and where $d_c(y)$ represents the distance between y and center \bar{y}_c in Y . For the power q used in (28), we normally choose the value 2.

It is easy to show that the above-given definitions indeed imply compliance with the requirements for the membership functions. More particularly we observe that, irrespective the concrete class center locations \bar{x}_b and \bar{y}_c , the fuzzy classes A_b and A_c have the important property (12), i.e.,

$$\forall x_p : \sum_{A_b} \mu_{A_b}(x_p) = 1 \quad \text{and} \quad \forall y_p : \sum_{A_c} \mu_{A_c}(y_p) = 1. \quad (29)$$

As a consequence of this, the probability density functions are functions on ‘well-defined’ fuzzy sample spaces X and Y , in the sense as explained in subsections 2.1.3 and 2.2.3.

3.3 The CELA-algorithm

We next re-formulate CELA within the general framework introduced in the previous section. For much more details than given below, we refer to [1, 2].

3.3.1 Step 1: Fixing the output classes

In order to get a concise characterization of the output space, the output classes A_c should be fixed by calculating appropriate locations of all fuzzy class centers in Y . This is done in the first step of CELA using a fuzzy clustering heuristic. Actually, we identify each fuzzy class A_c here with a fuzzy cluster and, in line with that, each class center \bar{y}_c with a cluster centroid. In the original approach, competitive learning was applied but other clustering algorithms might be useful as well.

The final result of the clustering is a fuzzy partitioning of Y such that each class center \bar{y}_c of the fuzzy class A_c is situated in the center of a ‘cloud’ of fuzzy sample points y_p . This also fixes the membership functions $\mu_{A_c}(y)$ describing the output classes A_c .

3.3.2 Step 2: Assessing the unconditional output probability distribution

Since the sample space Y is well-defined, the probability of each fuzzy class can be assessed properly by means of the approach introduced in subsection 2.3.1. To find the probability of the fuzzy classes A_c , the membership values of all data points y_p with respect to these fuzzy classes are summed up conform equation (25). This yields the probability vector

$$(\tilde{f}_{A_1}, \tilde{f}_{A_2}, \dots, \tilde{f}_{A_C}), \quad \text{with } \forall A_c : \tilde{f}_{A_c} = \frac{1}{P} \sum_{p=1}^P \mu_{A_c}(x_p) = \hat{\mu}_{A_c} \approx \Pr(A_c). \quad (30)$$

The probability vector $(\tilde{f}_{A_1}, \tilde{f}_{A_2}, \dots, \tilde{f}_{A_C})$, which can also be written as $(\hat{\mu}_{A_1}, \hat{\mu}_{A_2}, \dots, \hat{\mu}_{A_C})$, characterizes the unconditional behaviour in the output space. In [1], this approximation was called the ‘unconditional output cluster membership distribution’ (UOD).

3.3.3 Step 3: Fixing the input classes

In order to make an assessment of the conditional output distributions, the class centers \bar{x}_b in the input space should be fixed first. It is the most tricky part of CELA. Again a (fuzzy) competitive learning heuristic is used (although other approaches are conceivable). We only sketch the essentials concentrating on aspects related to the mathematical framework introduced. For every fuzzy sample point x_p , we calculate the ‘winning cluster’ A_b in X indicating which of the input clusters contributes at most to an exceptional sample point y_p . The ‘output exception’ $OE(p)$ of the corresponding output sample y_p is quantified using a ‘membership distance measure’ (something like a fuzzy Hamming distance) between the vector of membership values $(\mu_{A_1}(y_p), \mu_{A_2}(y_p), \dots, \mu_{A_C}(y_p))$ of y_p and the vector of average membership values $(\hat{\mu}_{A_1}, \hat{\mu}_{A_2}, \dots, \hat{\mu}_{A_C})$ as defined by equation (30):

$$OE(p) = \sqrt{\sum_{c=1}^{C_y} (\mu_{A_c}(y_p) - \hat{\mu}_{A_c})^2}. \quad (31)$$

The competitive learning heuristic used, maximizes an ‘exception fitness function’ (based on a summation of ‘winning’ output exceptions) by changing the class center locations \bar{x}_b in X . The net effect of it is that all class centers \bar{x}_b are fixed near those input clusters (of samples x_p) which correspond to the clusters of most exceptional data points y_p .

So after having performed this step, all classes A_b in the input space are fixed as well through the determination of the class center values of the corresponding membership functions (29).

3.3.4 Step 4: Assessing the conditional probability distributions

Using the given set of fuzzy sample pairs (x_p, y_p) and applying *conditional* mathematical statistics for fuzzy sets, we can calculate the conditional probabilities $\Pr(A_c|A_b)$ conform equation (27).

For each ‘regime’ A_b , ($b = 1, \dots, B$), the vector $(\Pr(A_1|A_b), \Pr(A_2|A_b), \dots, \Pr(A_C|A_b))$ gives a statistical characterization of exceptional conditional behaviour in the output space. In [1], we termed such a characterization a ‘conditional output cluster membership distribution’ (COD). Finally, we have found B exceptions, each one conditional on a regime described by one fuzzy event A_b .

3.3.5 Step 5: Deriving a fuzzy rule base

By comparing the various CODs found to the UOD as found in step 2 of CELA, we can determine the most exceptional relationships, i.e., the regimes for which the deviations from the UOD are most exceptional. The fuzzy distance measure as applied in (31), is useful here. Next, we try to express the most exceptional fuzzy relationships in a fuzzy rule base. For each regime, the deviations from the UOD can be expressed in linguistic terms. It could be helpful to exploit knowledge from experts working in the domain where CELA is applied. For a nice example on how such a fuzzy rule base is set up, we refer to [8].

4 Conclusions and future research

Let us first draw some conclusions on what we have achieved in this paper. The mathematical framework introduced in section 2, shows how probability theory and mathematical statistics can be combined with fuzzy set theory in order to introduce a clear mathematical framework. In domains where sample spaces should be defined using fuzzy sets and where, in addition, a ‘well-defined’ (in the sense as introduced) probability distribution is needed to describe the probabilistic fuzzy events of that domain, this framework might be helpful. A simple example shows how natural the framework can be applied to make a model of reality: calculating ‘the probability that a Dutch adult man is long’. If we choose randomly an arbitrary adult man, probability is clearly part of the modeling. In addition, a fuzzy set can be used to model the notion of ‘long’. Having a representative sample of P Dutch adult men, an assessment of the probability desired can rapidly be made using equation (25).

It has also been shown how the CELA-algorithm beautifully fits in this framework. Using notions like fuzzy classes and (un)conditional probability distributions, we get a deep understanding of how CELA actually works and also clarifies a lot of the assumptions behind this algorithm. For example, it has become clear now that we have indeed to do with an assessment of probability distributions defined on fuzzy sample spaces.

In future research, we want to continue along this line of research. E.g., it is interesting to analyze how certain choices as made in the implementation of CELA (like the clustering heuristics, the number of fuzzy classes chosen, and the general form of the membership functions) affect its performance. This can be analyzed both experimentally and mathematically. At the same time, we should continue to apply CELA in various domains. This can also help to get a better feeling of and insight in what this interesting algorithm is actually able to do. The insights obtained may in turn yield inspiration for the theoretical work to be done.

References

- [1] W.M. van den Bergh and J. van den Berg, “Competitive exception learning using fuzzy frequency distributions,” *Neural Network world*, vol. 10, no. 1-2, pp. 59-71, 2000.

- [2] J. van den Berg and W.M. van den Bergh, "On competitive exception learning," in *Proceedings of the 19th NAFIPS International Conference*, Atlanta, USA, 2000, pp. 34–38.
- [3] L.A. Zadeh, "Probability measures and fuzzy events," in *Fuzzy Sets and Applications, Selected Papers by L.A. Zadeh*, R.R. Yager, S. Ovchinnikov, R.M Tong, and H.T. Nguyen, Eds., pp. 45–51. John Wiley and Sons, USA, 1987.
- [4] W. Feller, *An Introduction to Probability Theory and its Applications*, John Wiley and Sons, USA, 1950.
- [5] R.V. Hogg and A.T. Craig, *Introduction to Mathematical Statistics*, MacMillan, London, 1970.
- [6] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, Upper Saddle River, 1995.
- [7] J.C. Dunn, "A fuzzy relative of the isodata process and its use in detecting compact, well-separated clusters," *Journal of Cybernetics*, vol. 3, no. 3, pp. 32–57, 1973.
- [8] Willem-Max van den Bergh, Jan van den Berg, and Uzay Kaymak, "Detecting noise trading using fuzzy exception learning," To appear in *Proceedings of IFSA/NAFIPS2001*, July 2001.

Publications in the Report Series Research* in Management

ERIM Research Program: "Business Processes, Logistics and Information Systems"

2001

Bankruptcy Prediction with Rough Sets

Jan C. Bioch & Viara Popova

ERS-2001-11-LIS

Neural Networks for Target Selection in Direct Marketing

Rob Potharst, Uzay Kaymak & Wim Pijls

ERS-2001-14-LIS

An Inventory Model with Dependent Product Demands and Returns

Gudrun P. Kiesmüller & Erwin van der Laan

ERS-2001-16-LIS

Weighted Constraints in Fuzzy Optimization

U. Kaymak & J.M. Sousa

ERS-2001-19-LIS

Minimum Vehicle Fleet Size at a Container Terminal

Iris F.A. Vis, René de Koster & Martin W.P. Savelsbergh

ERS-2001-24-LIS

The algorithmic complexity of modular decomposition

Jan C. Bioch

ERS-2001-30-LIS

A Dynamic Approach to Vehicle Scheduling

Dennis Huisman, Richard Freling & Albert Wagelmans

ERS-2001-35-LIS

Effective Algorithms for Integrated Scheduling of Handling Equipment at Automated Container Terminals

Patrick J.M. Meersmans & Albert Wagelmans

ERS-2001-36-LIS

Rostering at a Dutch Security Firm

Richard Freling, Nanda Piersma, Albert P.M. Wagelmans & Arjen van de Wetering

ERS-2001-37-LIS

Probabilistic and Statistical Fuzzy Set Foundations of Competitive Exception Learning

J. van den Berg, W.M. van den Bergh, U. Kaymak

ERS-2001-40-LIS

* A complete overview of the ERIM Report Series Research in Management:
<http://www.ers.erim.eur.nl>

ERIM Research Programs:
LIS Business Processes, Logistics and Information Systems
ORG Organizing for Performance
MKT Marketing
F&A Finance and Accounting
STR Strategy and Entrepreneurship

2000

A Greedy Heuristic for a Three-Level Multi-Period Single-Sourcing Problem

H. Edwin Romeijn & Dolores Romero Morales

ERS-2000-04-LIS

Integer Constraints for Train Series Connections

Rob A. Zuidwijk & Leo G. Kroon

ERS-2000-05-LIS

Competitive Exception Learning Using Fuzzy Frequency Distribution

W-M. van den Bergh & J. van den Berg

ERS-2000-06-LIS

Models and Algorithms for Integration of Vehicle and Crew Scheduling

Richard Freling, Dennis Huisman & Albert P.M. Wagelmans

ERS-2000-14-LIS

Managing Knowledge in a Distributed Decision Making Context: The Way Forward for Decision Support Systems

Sajda Qureshi & Vlatka Hlupic

ERS-2000-16-LIS

Adaptiveness in Virtual Teams: Organisational Challenges and Research Direction

Sajda Qureshi & Doug Vogel

ERS-2000-20-LIS

Assessment of Sustainable Development: a Novel Approach using Fuzzy Set Theory

A.M.G. Cornelissen, J. van den Berg, W.J. Koops, M. Grossman & H.M.J. Udo

ERS-2000-23-LIS

Applying an Integrated Approach to Vehicle and Crew Scheduling in Practice

Richard Freling, Dennis Huisman & Albert P.M. Wagelmans

ERS-2000-31-LIS

An NPV and AC analysis of a stochastic inventory system with joint manufacturing and remanufacturing

Erwin van der Laan

ERS-2000-38-LIS

Generalizing Refinement Operators to Learn Prenex Conjunctive Normal Forms

Shan-Hwei Nienhuys-Cheng, Wim Van Laer, Jan Ramon & Luc De Raedt

ERS-2000-39-LIS

Classification and Target Group Selection bases upon Frequent Patterns

Wim Pijls & Rob Potharst

ERS-2000-40-LIS

Average Costs versus Net Present Value: a Comparison for Multi-Source Inventory Models

Erwin van der Laan & Ruud Teunter

ERS-2000-47-LIS

Fuzzy Modeling of Client Preference in Data-Rich Marketing Environments

Magne Setnes & Uzay Kaymak

ERS-2000-49-LIS

Extended Fuzzy Clustering Algorithms

Uzay Kaymak & Magne Setnes

ERS-2000-51-LIS

Mining frequent itemsets in memory-resident databases

Wim Pijls & Jan C. Bioch

ERS-2000-53-LIS

Crew Scheduling for Netherlands Railways. "Destination: Customer"

Leo Kroon & Matteo Fischetti

ERS-2000-56-LIS