Fuzzy Sets in Data Analysis: Between Theory and Applications

Prof.Dr. Horia F. POP

E-mail: hfpop@cs.ubbcluj.ro
Web page: http://www.cs.ubbcluj.ro/~hfpop

Faculty of Mathematics and Computer Science, Babeş-Bolyai University

Cluj-Napoca, December 7, 2004

Motto:

"The fuzzy set was conceived as a result of an attempt to come to grips with the problem of pattern recognition in the context of imprecisely defined categories. In such cases, the belonging of an object to a class is a matter of degree, as is the question of whether or not a group of objects form a cluster."

The founder of Fuzzy Logic, Prof. Lotfi A. Zadeh, University of Berkeley (1981)

Contents

Fuzzy Sets – Definitions

- Fuzzy set $A : X \to [0,1]$, $A(x)$ is membership degree
- Special sets: $X(x) = 1$; $\emptyset(x) = 0$, $\forall x \in X$
- $A = B \iff A(x) = B(x), \forall x \in X$
- $A \subseteq B \iff A(x) \leq B(x), \forall x \in X$
- $\bar{A}(x) = 1 - A(x), \forall x \in X$
- $(A \cap B)(x) = T(A(x), B(x))$, T – triangular norm
- $(A \cup B)(x) = S(A(x), B(x))$, S – triangular conorm
- T, S commutative, associative, non-decreasing, $T(a,1) = a$, $S(a,0) = a$

Fuzzy Sets – Properties

- $T_S(a,b) = \min\{a,b\}$, $S_S(a,b) = \max\{a,b\}$
- $T_L(a,b) = \max\{a+b-1,0\}$, $S_L(a,b) = \min\{a+b,1\}$

	Property	Crisp sets	T_S and S_S	T_L and S_L
(1)	Idempotence laws	Valid	Valid	Invalid
(2)	$A \cup \bar{A} = X$, $A \cap \bar{A} = \emptyset$	Valid	Invalid	Valid
(3)	Distributivity laws	Valid	Valid	Invalid
(4)	De Morgan laws	Valid	Valid	Valid

Rough Sets – Alternative

- Approximation of sets using a collection of sets – Pawlak (1982)
- Given a collection of sets $C = \{C_1, C_2, \ldots\}$, and a set D
- *Lower approximation of D by C*: $D_L = \cup C_i$ such that $C_i \cap D = C_i$
- *Upper approximation of D by C*: $D^U = \cup C_i$ such that $C_i \cap D \neq \emptyset$
- *Boundary of D by C*: $D_L^U = D^U - D_L$
- A set D is *rough* if it has a non-empty boundary when approximated by C
- Otherwise, the set D is *crisp*

- Fuzzy rough sets, rough fuzzy sets

Fuzzy clustering – Discussion

Data representation: objects are vectors of measured values

Feature extraction: to reduce data dimensionality and eliminate redundant characteristics

Clusters shape: different geometric prototypes; norms or scalar products

Clusters size: use of adaptive distance or adaptive algorithms

Clusters validity: optimal number of classes through validity functionals, clusters merging/splitting or by using a hierarchical approach

Final fuzzy partition: needs to be defuzzified; it should not be discarded

Method: fuzzy objective function minimization; two step iterative procedure that continually decreases the value of the objective function

Fuzzy clustering – Generic algorithm

Aim: minimize $J(P, L) = \sum_{i=1}^{c} \sum_{j=1}^{n} A_i(x^j)^m \cdot D(x^j, L_i)$

1. Given: c, n, m, and $x^j, j = 1, \ldots, n$; $l = 0$;
2. Initialize fuzzy partition $P^{(0)} = \{A_1, \ldots, A_c\}$
3. Compute prototypes L_i that minimize $J(P^{(l)}, \cdot)$ – may be costly
4. Compute fuzzy partition $P^{(l+1)}$ that minimizes $J(\cdot, L)$
$$A_i^{(l+1)}(x^j) = \frac{1}{\sum_{k=1}^{c} \left(\frac{D(x^j, L_i)}{D(x^j, L_k)} \right)^{\frac{1}{m-1}}}$$
5. Compare $P^{(l+1)}$ with $P^{(l)}$. If close enough, then STOP, else increase l by 1 and GOTO STEP 3

Fuzzy clustering – Variants and improvements

Geometric prototypes

- Fuzzy c-Means – Dunn (1974), Bezdek (1974)
- Fuzzy c-Varieties and Fuzzy c-Elliptotypes – Bezdek et.al. (1981)
- Fuzzy c-Ellipsoids – Lenart (1989)
- Adaptive Fuzzy Clustering – Dave (1989), Dumitrescu, Pop (1990)
- Use of fuzzy covariance matrix – Gustaffson, Kessel (1979)
- L_p Fuzzy c-Means – Miyamoto, Agusta (1998), Hathaway, Bezdek (2000)

Empty shell prototypes

- Fuzzy c-Shells – Dave (1990), Krishnapuram, Nasraoui, Frigui (1992)
- Adaptive Fuzzy c-Shells – Dave, Bhaswan (1992)
- Fuzzy c-Ellipsoidal Shells – Frigui, Krishnapuram (1996), Gath, Hoory (1995)
- Fuzzy c-Quadric Shells – Krishnapuram, Frigui, Nasraoui (1993, 1995)
- Fuzzy c-Rectangular Shells – Höppner, Klawonn, Kruse (1997)

Fuzzy clustering – Variants and improvements (2)

Fuzzy clustering with incomplete data

- Unsupervised fuzzy competitive learning – Chung, Lee (1994)
- Whole data strategy – Hathaway, Bezdek (2001)
- Partial distance strategy – Hathaway, Bezdek (2001)
- Optimal completion strategy – Hathaway, Bezdek (2001)
- Nearest neighbour strategy – Hathaway, Bezdek (2001)

Other methods and models

- Fuzzy divisive hierarchic clustering – Dumitrescu (1988)
- Cross-clustering – Dumitrescu, Pop (1995)
- Noise Clustering – Dave (1991), Dave, Sen (1997)
- Possibilistic, Probabilistic – Krishnapuram, Keller (1993), Gath, Geva (1989)

Fuzzy classification

Crisp input, crisp output

- The original approach

Fuzzy input, crisp output

- Most fuzzy generalisations

Fuzzy input, fuzzy output

- Fuzzy Nearest k Neighbors
- Fuzzy Nearest Prototypes
- Restricted Fuzzy Clustering – Pop (1995)
- Fuzzy Decision Hyperplanes – Lenart (1993)
- Fuzzy Competitive Learning Variants

Fuzzy classification – Details

Restricted Fuzzy Clustering

- Consider the set X, fuzzy partition P, extra item x_0
- Use Fuzzy Clustering on the set $X \cup \{x_0\}$ but freeze fuzzy membership degrees of data items in X

Fuzzy Competitive Learning

1. v_k^i initialized as random; iteration $= 1$; step $= 1.0$;
2. loop for all data items x^j
   - loop for all k: if x_k^j is available then, for all i,

$$v_k^i = v_k^i + \text{step} \cdot A_i(x^j)^m \cdot (x_k^j - v_k^i)$$

3. iteration $=$ iteration $+ 1$; step $=$ step/iteration
4. if old and new centers v are close enough, then stop, else goto step 2

Regression techniques – The problem

Aim: to relate, corelate or model a measure response based on the value of a given variable

Common approach: use the linear Least-Square method

Assumption: data is homoscedastic – y-direction error is independent of the controlled variable

Robustness: heteroscedastic data and presence of outliers are common

Problem: most of current robust methods do not provide best results in all situations

Problem: quality measures are not method independent

Suggestion: need for fuzziness as a way to model data items weights

Validation: direct comparative analysis, cross-validation, etc

Fuzzy regression – Discussion

Clustering techniques: able to detect the cluster substructure of a data set; do not work for $c = 1$

Necessity: to determine the fuzzy set A and the prototype L that best describes the data set

Motivation: useful in search of good robust regression methods: the weights are fuzzy membership degrees

Method: determine a fuzzy partition $\{A, \bar{A}\}$; $\bar{A}$ is a class with a hypothetical prototype, characterized by $D(x^j, \bar{L}) = \delta := \left(\frac{\alpha}{1-\alpha}\right)^{m-1}$

Objective function:

$$J(A, L) = \sum_{j=1}^{n} A(x^j)^m D(x^j, L_i) + \sum_{j=1}^{n} \bar{A}(x^j)^m \left(\frac{\alpha}{1-\alpha}\right)^{m-1}, \ \alpha \in (0,1)$$

Fuzzy regression – Generic algorithm

1. Given α; Initialize $A^{(0)}(x) = 1$, $l = 0$
2. Compute prototype L that minimizes $J(A^{(l)}, \cdot)$
3. Compute fuzzy set $A^{(l+1)}$ that minimizes $J(\cdot, L)$

$$A^{(l+1)}(x^j) = \frac{\dfrac{\alpha}{1-\alpha}}{\dfrac{\alpha}{1-\alpha} + D(x^j, L)^{\frac{1}{m-1}}}$$

4. Compare $A^{(l+1)}$ with $A^{(l)}$. If close enough, then STOP, else increase l by 1 and GOTO STEP 3

Improvement: scale independence: introduction in step 3 of relative distances

Fuzzy regression – Properties

i. $A(x) = 1 \Leftrightarrow D(x, L) = 0$ $\quad$ $A(x) = \alpha \Leftrightarrow D_r(x, L) = 1$

ii. $A(x) \in [\alpha, 1]$ for all $x \in X$ $\quad$ (consequence)

iii. $\alpha = 0 \Leftrightarrow A(x) = 0$ for all $x \in X$ $\quad$ $\alpha = 1 \Leftrightarrow A(x) = 1$ for all $x \in X$

iv. $A(x) = A(y) \Leftrightarrow D(x, L) = D(y, L)$ $\quad$ $A(x) < A(y) \Leftrightarrow D(x, L) < D(y, L)$

- constant α sets the polarization of the fuzzy partition $\{A, \bar{A}\}$
- Remarkable flexibility; best results with $\alpha \approx 0.10$
- Linear version (Fuzzy Linear Regression)

(a) dissimilarity: square distance to the line (not vertical distance)

(b) linear prototype determined by fuzzy mean and principal component

(c) operational in all testing conditions; better than most methods

(d) allows detection of data type through repeated runs

Robust regression – Other approaches

Non-fuzzy regression

- Weighted Least-Square
- Iterative Reweighted Least-Square
  (weights - inverse of variance, residuals, etc)

Fuzzy regression

- Deal with fuzzy and crisp data; use of fuzzy regression coefficients
- Using minimium fuzziness criterion
- Using least-squares of errors criterion
- Using interval analysis (data and coefficients are fuzzy intervals)
  (Chang, Ayyub, 2001)

Data projection and selection

- Visualisation of high-dimensional data items
- Projection methods
  - Linear projection methods (e.g. PCA)
  - Non-linear projection methods (e.g. MDS)
- Variables selection methods

Data analysis methods

- use of projection methods to reduce data dimensionality for further clustering
  - certain projection methods are better for to be used together with certain clustering methods
- use of projection methods as a visualisation tool, to help in understanding the clustering results
  - display the reference vectors using some distance-preserving projection method
- use of MDS methods as a way to apply metric-based clustering algorithms to non-metric data
  - create a metric space with a distance that best preserves the original dissimilarities

Principal Components Analysis (PCA)

- Aim: data dimensionality reduction by determining new, fewer variables
- The new variables are called principal components and correspond to the axes of maximal elongation of data
- The number of principal components necessary to conserve 90% of data variance is considerably less than the size of data space
- The principal components are linear combinations of the original variables; need to determine the relevant variables
- Problem: Points isolated with respect to (a) the data set; (b) principal directions

PCA Algorithm

1. Compute the covariance or corelation matrix

$$\text{Cov}_{ij} = \frac{1}{n-1} \sum_{k=1}^{n} (x_i^k - \bar{x}_i) \cdot (x_j^k - \bar{x}_j)$$

$$\text{Cor}_{ij} = \frac{\text{Cov}_{ij}}{s_i \cdot s_j}, s_i = \frac{1}{n-1} \sum_{k=1}^{n} (x_i^k - \bar{x}_i)^2$$

2. Compute the eigenvectors and eigenvalues of this matrix; these are the principal components and the scatter values

3. Based on these scatter values, select the necessary number of principal components

4. Determine the values of data for the new variables (i.e. project the data set in the space of the selected principal components): $X'^T = X^T \cdot E$

Fuzzy PCA, first component

- Problem: points isolated with respect to the first component ONLY
- Method: Membership degrees according to the distance to the first component
- Fuzzy Regression with linear prototype = eigenvector corresponding to the largest eigenvalue of the fuzzy covariance matrix

$$C_{ij} = \frac{\sum_{k=1}^{n} A(x^k)^m \cdot (x_i^k - \bar{x}_i) \cdot (x_j^k - \bar{x}_j)}{\sum_{k=1}^{n} A(x^k)^m}, \quad i, j = 1, \ldots, p. \tag{1}$$

- Advantage: the first principal component will count the merits of each data item; as such, will consider the isolated points with less significance
- To be optimised: the other principal components

Fuzzy PCA, orthogonal

- Problem: fuzzyfication of all components, not only the first one
- Idea: A different approach, by projecting the data in smaller-sized spaces
- After the first fuzzy eigenvector is determined, all data is projected to the hyperplane rectangular on it
- The eigenvectors corresponding to the projected data will be orthogonal to the eigenvector determined above
- The projection continues further on, etc.; finally, the eigenvectors are rebuilt in the original space
- Advantage: the compotation of the other fuzzy components is reduced to the computation of the first fuzzy component of a smaller-sized matrix

FuzzyPCAFirst Algorithm

1. Determine the optimal α optimal (leads to maximisation of a quality criterion of the first principal component)
2. Determine the fuzzy membership degrees for the α determined above
3. Using these fuzzy membership degrees, compute the fuzzy covariance matrix C (1)
4. Compute the eigenvectors and eigenvalues of C; these are the fuzzy principal components and the coresponding scatter values

FuzzyPCAOrthogonal Algorithm

1. If $p \leq 2$, use $\textsc{FuzzyPCAFirst}()$; otherwise, continue
2. Determine the optimal α optimal (leads to maximisation of a quality criterion of the first principal component)
3. Determine the fuzzy membership degrees for the α determined above
4. Using these fuzzy membership degrees, compute the fuzzy covariance matrix C (1)
5. Compute the eigenvectors and eigenvalues of C; the maximal eigenvalue λ and its eigenvector e are called 'current'
6. Compute the data scores and remove the values on the first positions
7. Recursively call $\textsc{FuzzyPCAOrthogonal}()$ on this set of reduced-size and determine the eigenvalues and eigenvectors from this projected space
8. Return to the original space and rewrite these eigenvectors and eigenvalues in terms of the original set of coordinates. The new values are $\lambda_2, \ldots, \lambda_p$ and $e^2, \ldots, e^p$

Multidimensional scaling (MDS)

- usefullness with dimensionality reduction and non-metric data
- metric MDS: $E_M = \sum_{k \neq l} (d_{kl} - d'_{kl})^2$
- non-metric MDS: $E_N = \sum_{k \neq l} (f(d_{kl}) - d'_{kl})^2 / \sum_{k \neq l} (d'_{kl})^2$
  - f – monotonically increasing; maps original distances to such values that best preserve the rank order
- Sammon mapping: $E_S = \sum_{k \neq l} (d_{kl} - d'_{kl})^2 / d_{kl}$
- $d_{kl} = d(k, l)$ is the dissimilarity of original items x_k and x_l
- $d'_{kl} = d'(k, l)$ is the corresponding distance in the projected space.

Software: SADIC

System for Automatic Data Investigation and Classification

Implementation:

- C++: object oriented coding
- wxWidgets: platform-independent class library
- Any C++ compiler on Windows, Linux and MacOS
- Console application with command line interpreter
- Visual application
- XML reports; XSL transformation to HTML output
- Easy to extend; no redundant code

Methods:

- Dimensionality reduction: (F)PCA, MDS
- Selection of relevant variables
- (Fuzzy) solid clustering
- (Fuzzy) incomplete data clustering
- (Fuzzy) shell clustering
- Horizontal, hierarchic and cross-clustering
- (Fuzzy) supervised classification
- (Fuzzy) (non-)linear regression
- ...

Applications of Fuzzy Data Analysis

- Optimal selection of solvent systems
- Roman pottery (terra sigillata)
- Greek muds and pelloids
- Fuzzy system of chemical elements
- Romanian and American coals
- Intramolecular interactions and catalyst modeling
- Assesment of heart disease
- Electric network distribution systems
- . . .

Final word

- Prof. Dan Dumitrescu
- Prof. Militon Frențiu
- Prof. Costel Sârbu, Prof. Thomas Cundari

- Prof. Lotfi Zadeh (founder of Fuzzy sets)
- Prof. Donald E. Knuth, Leslie Lamport (creators of TEX and LATEX)

- Cluj C.S. Department and Faculty members
- ...

"[. . .] When I read (and reread) their papers and discuss them with my students, I am often struck by the whimsical image of Mendeleev sitting at a computer, entering his collected elemental properties into the fuzzy classification programs developed by the Cluj team, and assessing the various elemental relationships at different levels of fuzziness! How much sooner would eka-aluminum and other new elements have been isolated had Mendeleev had such tools at his fingertips? [. . .]"

Prof. Thomas Cundari, University of Memphis (2000)