Streamlining the Applied Mathematics Studies
at Faculty of Science of Palacký University in Olomouc
CZ.1.07/2.2.00/15.0243

evropský
sociální
fond v ČR

EVROPSKÁ UNIE

MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

OP Vzdělávání
pro konkurenceschopnost

INVESTICE
DO ROZVOJE
VZDĚLÁVÁNÍ

International Conference
Olomoucian Days of Applied Mathematics

ODAM 2013

Department of Mathematical analysis
and Applications of Mathematics
Faculty of Science
Palacký Univerzity Olomouc

Fuzzy Sets and Rough Sets in Prototype- based Clustering Algorithms

Slavka Bodjanova
Texas A&M University-Kingsville,
Kingsville, TX 78363

ODAM
Olomouc, June 12, 2013

A brief introduction to my lecture

that will be posted on the ODAM web page.....

”So far as laws of mathematics refer to reality, they are not certain. And so far as they are certain, they do not refer to reality.”
(Albert Einstein, 1921)

”So far as laws of mathematics refer to reality, they are not certain. And so far as they are certain, they do not refer to reality.”
(Albert Einstein, 1921)

”The central problem of our age is how to act decisively in the absence of certainty.”
(Bertrand Russell, 1940)

Outline

1. Clustering and clustering algorithms
2. Fuzzy sets in c-means clustering
3. Rough sets in c-means clustering
4. Rough-fuzzy c-means clustering algorithms
5. Shadowed sets in rough-fuzzy c-means clustering

Clustering: standard approach in data mining searching for natural groups (clusters) present in data

Cluster: collection of objects similar to each other and dissimilar to objects from other clusters

Prototype-based clustering algorithms:

Prototype-based clustering algorithms:

- the number of clusters is predefined,

Prototype-based clustering algorithms:

- the number of clusters is predefined,
- each cluster is represented by its prototype,

Prototype-based clustering algorithms:

- the number of clusters is predefined,
- each cluster is represented by its prototype,
- objects are partitioned into clusters based on their similarity to clusters' prototypes by optimizing an **objective function**,

Prototype-based clustering algorithms:

- the number of clusters is predefined,
- each cluster is represented by its prototype,
- objects are partitioned into clusters based on their similarity to clusters' prototypes by optimizing an **objective function**,
- **c-means algorithm:** iterative algorithm which represents each cluster by its center of gravity (the mean).

Notation:

$X = \{x_1, \ldots, x_n\},\ x_j \in \mathbb{R}^p,\ j = 1, \ldots, n$ (set of objects)
$U = \{U_1, \ldots, U_c\},\ U_i \subset X$ (c-partition of X)
$U_i(x_j) = u_{ij}$ (membership of object x_j in cluster U_i)

$v_i \in \mathbb{R}^p$ (mean of cluster U_i)
$d(x_j, v_i) = ||x_j - v_i|| = d_{ij}$ (distance between x_j and v_i)

Hard c-means algorithm (HCM)
(k-means algorithm)

J.B. MacQueen (1967)

Minimizes the objective function $J = \sum_{j=1}^{n} \sum_{i=1}^{c} d_{ij}^2$.

Hard c-means algorithm (HCM)
(k-means algorithm)

J.B. MacQueen (1967)

Minimizes the objective function $J = \sum_{j=1}^{n} \sum_{i=1}^{c} d_{ij}^2$.

The main steps:

1) Assign initial means v_i.
2) Assign each x_j to the cluster U_i with the closest mean.
3) Compute the new mean

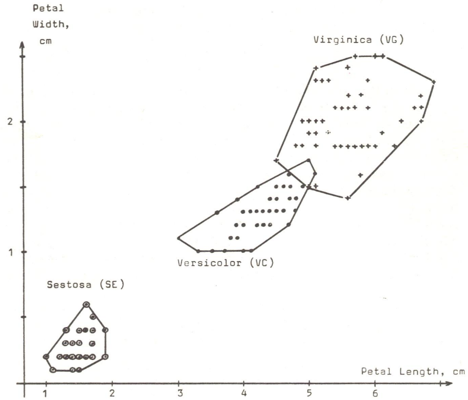
$$v_i = \frac{\sum_{x_j \in U_i} x_j}{|U_i|}.$$

Repeat steps 2) and 3) until there are no more new assignments of objects.

Hard c-means clustering algorithm (HCM) is an appropriate method of clustering when the analyzed data consists of

well separated clusters.

Anderson's iris data (petal featurs)



J.C. Bezdek: Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.

2. Fuzzy sets in c-means clustering

Lotfi Zadeh (1965)

Fuzzy sets: based on multivalued logic (partial truth, degree of truth) (Jan Lukasiewicz (1930), Max Black (1937))

2. Fuzzy sets in c-means clustering

Lotfi Zadeh (1965)

Fuzzy sets: based on multivalued logic (partial truth, degree of truth) (Jan Lukasiewicz (1930), Max Black (1937))

When A is a fuzzy subset of an universal set X and x is an object in X then the proposition

"x is a member of A"

is **true to the degree** to which x is actually representing the concept described by A.

2. Fuzzy sets in c-means clustering

Lotfi Zadeh (1965)

Fuzzy sets: based on multivalued logic (partial truth, degree of truth) (Jan Lukasiewicz (1930), Max Black (1937))

When A is a fuzzy subset of an universal set X and x is an object in X then the proposition

"x is a member of A"

is **true to the degree** to which x is actually representing the concept described by A.

Mathematically: membership function of A

$$\mu_A : X \rightarrow [0, 1].$$

Simplified notation: $\mu_A(x) = A(x)$.

Fuzzy c-means algorithm (FCM)

J.C. Bezdek (1981)
Fuzzification of the hard c-means algorithm:

$$u_{ij} \in [0, 1]$$

for all $u_{ij} \in U$.

Fuzzy c-means algorithm (FCM)

J.C. Bezdek (1981)
Fuzzification of the hard c-means algorithm:

$$u_{ij} \in [0, 1]$$

for all $u_{ij} \in U$.

Minimizes the objective function

$$J = \sum_{i=1}^{n} \sum_{i=1}^{c} (u_{ij})^m d_{ij}^2,$$

subject to

$$\sum_{i=1}^{c} u_{ij} = 1, \text{ for all } j,$$

where $1 \leq m < \infty$ is the fuzzifier.

Evaluation of membership grade u_{ij}:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}}.$$

Evaluation of membership grade u_{ij}:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}}.$$

Evaluation of centroid v_i:

$$v_i = \frac{\sum_{j=1}^{n} (u_{ij})^m x_j}{\sum_{j=1}^{n} (u_{ij})^m}.$$

Evaluation of membership grade u_{ij}:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}}.$$

Evaluation of centroid v_i:

$$v_i = \frac{\sum_{j=1}^{n} (u_{ij})^m x_j}{\sum_{j=1}^{n} (u_{ij})^m}.$$

Coefficient u_{ij} can be interpreted as the (posterior) probability $p(i/x_j)$ that, given x_j, it came from class i.

FCM is sensitive to noise and outliers.

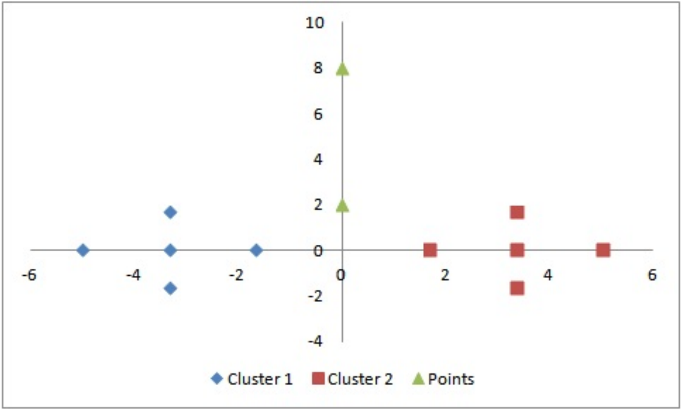
In FCM, the constraint $\sum_{i=1}^{c} u_{ij} = 1$ is too strong.

If x_j is equidistant from two different means v_r and v_s then

$$u_{rj} = u_{sj} = 1/2,$$

regardless whether the actual distance is large or small.

Assignment of points to clusters



Possibilistic c-partition of X:

$0 < \sum_{i=1}^{c} u_{ij} < c$ for all $x_j \in X$.

Possibilistic c-means algorithm (PCM)
R. Krishnapuram, J.M. Keller (1993)

FPCM algorithm - combination of FCM and PCM
N.R.Pal, K.Pal, J.M. Keller, J.C. Bezdek (2005)

In real life situations, uncertainty may arise from

incompleteness in class definition.

3. Rough sets in c-means clustering

In real life situations, uncertainty may arise from

incompleteness in class definition.

This type of uncertainty can be handled by

rough set theory.

Zdislaw Pawlak (1982)

X - finite non-empty universal set
$R \subset X \times X$- equivalence relation
(X, R) - approximation space (knowledge base)

$[x]_R$- equivalence class of R containing $x \in X$

Given $S \subset X$, it may not be possible to describe S precisely in approximation space (X, R).
Instead, one may characterize S by a pair of lower and upper approximations

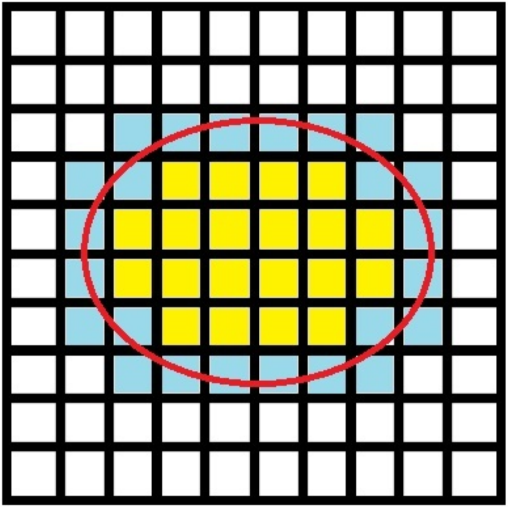
$$\underline{R}(S) = \{x \in X : [x]_R \subset S\},$$
$$\overline{R}(S) = \{x \in X : [x]_R \cap S \neq \emptyset\}.$$

Given $S \subset X$, it may not be possible to describe S precisely in approximation space (X, R).
Instead, one may characterize S by a pair of lower and upper approximations

$$\underline{R}(S) = \{x \in X : [x]_R \subset S\},$$
$$\overline{R}(S) = \{x \in X : [x]_R \cap S \neq \emptyset\}.$$

$(\underline{R}(S), \overline{R}(S))$ is a rough set with reference set S

Example of a rough set



Rough c-means algorithm (RCM)
(rough k-means algorithm)

P. Lingras and C. West (2004)

Cluster $U_i \in U$ is characterized by

lower approximation $\underline{A}(U_i)$ and upper approximation $\overline{A}(U_i)$.

(interval-based representation of rough sets)

Rough c-means algorithm (RCM)
(rough k-means algorithm)

P. Lingras and C. West (2004)

Cluster $U_i \in U$ is characterized by

lower approximation $\underline{A}(U_i)$ and upper approximation $\overline{A}(U_i)$.

(interval-based representation of rough sets)

Properties:

1) $x_j \in X$ can be part of at most one lower approximation,

Rough c-means algorithm (RCM)
(rough k-means algorithm)

P. Lingras and C. West (2004)

Cluster $U_i \in U$ is characterized by

lower approximation $\underline{A}(U_i)$ and upper approximation $\overline{A}(U_i)$.

(interval-based representation of rough sets)

Properties:

1) $x_j \in X$ can be part of at most one lower approximation,

2) if $x_j \in \underline{A}(U_i)$ then $x_j \in \overline{A}(U_i)$,

Rough c-means algorithm (RCM)
(rough k-means algorithm)

P. Lingras and C. West (2004)

Cluster $U_i \in U$ is characterized by

lower approximation $\underline{A}(U_i)$ and upper approximation $\overline{A}(U_i)$.

(interval-based representation of rough sets)

Properties:

1) $x_j \in X$ can be part of at most one lower approximation,

2) if $x_j \in \underline{A}(U_i)$ then $x_j \in \overline{A}(U_i)$,

3) if x_j is not a part of any lower approximation, then it belongs to two or more upper approximations.

The main steps:

1) Assign initial means $v_i, \ i = 1, \ldots, c$, choose threshold TH.

2) For each $x_j \in X$ compute its distance d_{ij} from the cluster centroides $v_i, \ i = 1, \ldots, c$.

3) Let $d_{rj} = \min\{d_{ij}, i = 1, \ldots, c\}$ and $\Omega = \{i : \frac{d_{ij}}{d_{rj}} \leq TH, i \neq r\}$.

If $\Omega = \emptyset$ then $x_j \in \underline{A}(U_r)$,
otherwise $x_j \in \overline{A}(U_r)$ and $x_j \in \overline{A}(U_i)$ for all $i \in \Omega$.

4) Compute new mean v_i for each cluster U_i.

Repeat steps 2) and 4) until there are no more new assignments of objects.

Threshold TH

If $TH \to 1$ then $\Omega \to \emptyset$ and RCM $\to$ HCM

Main area of application of RCM

when ”clear cases” need to be distinguished from ”unclear” (e.g., quality control: good products, products with some doubts).

Rough-fuzzy c-means algorithm (RFCM)

S. Mitra, H. Banka, W. Perycz (2004)

Both the lower and the upper approximations of a cluster are fuzzy sets.

Rough- fuzzy- possibilistic c-means algorithm (RFPCM)

Pradipta Maji and Sankar K. Pal (2007)

The lower approximation of a cluster is a crisp set, the upper approximation is a fuzzy set.

5. Shadowed sets in rough-fuzzy c-means clustering

In rough-fuzzy c-means algorithms proposed by Mitra et al. (2004) and Maji and Pal (2007),

the lower approximation and the boundary of each cluster U_i depend on a **fixed threshold** $\boldsymbol{TH}$.

In rough-fuzzy c-means algorithms proposed by Mitra et al. (2004) and Maji and Pal (2007),

the lower approximation and the boundary of each cluster U_i depend on a **fixed threshold** TH.

Modification:

Shadowed c-means algorithm (SCM)

S. Mitra, W. Pedrycz, B. Barman (2010)

SCM provides **dynamical evaluation of** TH_i for each cluster U_i based on available data.

W. Pedrycz (1998)

Shadowed set $\widehat{f}$ induced by a fuzzy set f on X
is an interval-valued set on X that maps elements $x \in X$ into 0, 1, and the interval $(0, 1)$ such that :

W. Pedrycz (1998)

Shadowed set $\widehat{f}$ induced by a fuzzy set f on X
is an interval-valued set on X that maps elements $x \in X$ into 0, 1, and the interval $(0, 1)$ such that :

$$\widehat{f}(x) = \begin{cases} 0 & \text{if } f(x) \leq \lambda, \\ (0,1) & \text{if } \lambda < f(x) < f_{\max} - \lambda, \\ 1 & \text{if } f(x) \geq f_{\max} - \lambda, \end{cases}$$

W. Pedrycz (1998)

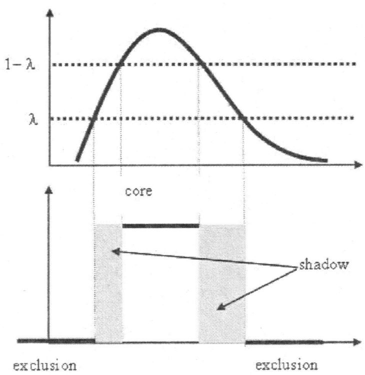
Shadowed set $\widehat{f}$ induced by a fuzzy set f on X is an interval-valued set on X that maps elements $x \in X$ into 0, 1, and the interval $(0, 1)$ such that :

$$\widehat{f}(x) = \begin{cases} 0 & \text{if } f(x) \leq \lambda, \\ (0,1) & \text{if } \lambda < f(x) < f_{\max} - \lambda, \\ 1 & \text{if } f(x) \geq f_{\max} - \lambda, \end{cases}$$

$f_{\max} = \max\{f(x), x \in X\}$ and

λ is derived from the membership grades of f such that the changes of membership grades to 0 or 1 are compensated for by the construction of the **“ shadow” represented by the interval (0, 1).**

Fuzzy set inducing a shadowed set via a threshold



S. Mitra, W. Pedrycz, B. Barman: Shadowed c-means: Integrating fuzzy and rough clustering. *Pattern Recognition* 43, (2010) 1282–1291.

Computation of TH_i for cluster U_i based on shadowed sets

Let $u_{imin} = \min\{u_{ij}, x_j \in X\}$ and $u_{imax} = \max\{u_{ij}, x_j \in X\}$.

Find $TH_i \in (u_{imin}, u_{imax})$ which minimizes

Computation of TH_i for cluster U_i based on shadowed sets

Let $u_{imin} = \min\{u_{ij}, x_j \in X\}$ and $u_{imax} = \max\{u_{ij}, x_j \in X\}$.

Find $TH_i \in (u_{imin}, u_{imax})$ which minimizes

$$V(TH_i) = \left| \sum_{x_j \in E} u_{ij} + \sum_{x_j \in C} (u_{imax} - u_{ij}) - |B| \right|,$$

where
$E = \{x_j \in X : u_{ij} \leq TH_i\}$ (exclusion region of U_i),
$B = \{x_j \in X : TH_i < u_{ij} < u_{imax} - TH_i\}$ (boundary of U_i = shadow),
$C = \{x_j \in X : u_{ij} \geq u_{imax} - TH_i\}$ (lower bound of U_i = core of U_i).

The main steps of Shadowed c-means algorithm:

1) Assign initial means v_i, choose fuzzifier m.
2) Compute u_{ij} as in FCM.

3) Compute threshold TH_i for each cluster U_i using shadowed sets.

4) Update means $v_i, \ i = 1, \ldots, c$.

Repeat steps 2) and 4) until there are no more new assignments of objects.

Computation of the mean v_i of cluster U_i based on shadowed sets

$$v_i = \frac{\sum_{x_j \in C} x_j + \sum_{x_j \in B} (u_{ij})^m x_j + \sum_{x_j \in E} (u_{ij})^{m^m} x_j}{|C| + |B| + |E|},$$

where m is the fuzzifier from FCM algorithm.

Computation of the mean v_i of cluster U_i based on shadowed sets

$$v_i = \frac{\sum_{x_j \in C} x_j + \sum_{x_j \in B} (u_{ij})^m x_j + \sum_{x_j \in E} (u_{ij})^{m^m} x_j}{|C| + |B| + |E|},$$

where m is the fuzzifier from FCM algorithm.

Assignment of weights:

1 when $x_j \in C$ (strong influence of elements from core),
u_{ij}^m when $x_j \in B$ (medium influence of elements from boundary),
$u_{ij}^{m^m}$ when $x_j \in E$ (low influence of elements from exclusion region).

Advantages of SCM:

more realistic modeling of data, better estimation of prototypes, robust in the presence of outliers.

Advantages of SCM:

more realistic modeling of data, better estimation of prototypes, robust in the presence of outliers.

SCM demonstrated that it is possible to filter out irrelevant information while remaining in fuzzy framework.

Advantages of SCM:

more realistic modeling of data, better estimation of prototypes, robust in the presence of outliers.

SCM demonstrated that it is possible to filter out irrelevant information while remaining in fuzzy framework.

”The three-valued quantification of the resulting cluster structure helps us easily identify regions (and patterns) which may require further attention while pointing at the core structure and patterns that arise with high values of typicality with respect to detected clusters.” (Mitra, Pedrycz, Barman, 2010)

Thank you for your attention.

References

- J.C. Bezdek: *Pattern Recognition with Fuzzy Objective Function Algorithm.* Plenum Press, New York (1981).
- R. Krishnapuram, J.M. Keller: A possibilistic approach to clustering. *IEEE Trans. Fuzzy Syst.* vol.1, no.2, (1993) 98–110.
- P. Lingras, C. West: Intervak set clustering of Web users with rough k-means. *Jpurnal of Intelligent Information Systems* 23, (2004) 5–16.
- J.B. MacQueen: Some methods for classificaiton and Analysis of multivariate observations. *1. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, (1967) 281–297.
- P. Maji, S. Pal: Rough ste based generalized fuzzy c-means algorithm and quantitative indices. *IEEE Trans. Syst., Man and Cybernetics* Part B, vol. 37, (2007) 1529–1540.

References

- S.Mitra, H. Banka, W. Pedrycz: Rough-fuzzy collaborative clustering. *IEEE Trans. Syst., Man and Cybernetics* Part B, vol. 37, (2005) 795–805.
- S. Mitra, W. Pedrycz, B. Barman: Shadowed c-means: Integrating fuzzy and rough clustering. *Pattern Recognition* 43, (2010) 1282–1291.
- N.R. Pal, K. Pal, J.M. Keller, J.C. Bezdek: A possibilistic fuzzy c-menas clustering algorithm. *IEEE Trans. Fuzzy Syst.* vol.13, no.4, (2005) 517–530.
- Pawlak, Z.: Rough Sets. *International Journal of Computer and Information Sciences* 11, (1982) 341–356.).
- Peters, G., Crespo, F., Lingras, P., Weber, R.: Soft clustering-fuzzy and rough approaches and their extensions and derivatives. *International Journal of Approximate Reasoning 54, (2013) 307–322.*
- *L.A. Zadeh: Fuzzy Sets.* Information and Control *8, (1965) 338–353.*