

AN INTUITIONISTIC FUZZY CLUSTERING TECHNIQUE



ТЕХНИКА ЗА ИНТУИЦИОНИСТКИ
РАЗМИТ КЛЪСТЕРИНГ

DR. R. PARVATHI
DEPT. OF MATHEMATICS, VELLALAR COLLEGE FOR WOMEN
ERODE, TAMILNADU, INDIA
paarvathis@rediffmail.com

OUTLINE

СЪДЪРЖАНИЕ

- Fuzzy and Intuitionistic Fuzzy Sets (IFSs)
Размити и интуиционистки размити множества
- Algorithms
Алгоритми
 - Clustering
Клъстеринг
 - Fuzzy Clustering
Размит клъстеринг
 - IF clustering technique
Техника за ИР клъстеринг
 - Distributed clustering
Разпределен клъстеринг
- Applications of Clustering
Приложения на клъстеринга

INTUITIONISTIC FUZZY SETS

ИНТУИЦИОНИСТКИ РАЗМИТИ МНОЖЕСТВА

- Crisp sets

Стандартни (булеви) множества

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in X \\ 0 & \text{if } x \notin X \end{cases}$$

that is

$$\chi_A : X \rightarrow \{0,1\}$$

- Fuzzy sets. Membership function

Размити множества. Функция на принадлежност

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

INTUITIONISTIC FUZZY SETS

ИНТУИЦИОНИСТКИ РАЗМИТИ МНОЖЕСТВА

- Intuitionistic fuzzy sets

Интуиционистки размити множества

Membership function

Функция на принадлежност

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

Non-membership function

Функция на непринадлежност

$$\nu_A : X \rightarrow [0,1]$$

so that:

така че:

$$0 \leq \mu_A(x) + \nu_A(x) \leq 1$$

INTUITIONISTIC FUZZY SETS

ИНТУИЦИОНИСТКИ РАЗМИТИ МНОЖЕСТВА

- Degree of uncertainty
Степен на неопределеност

$$\pi_A(x) = 1 - \mu_A(x) - \nu_A(x)$$

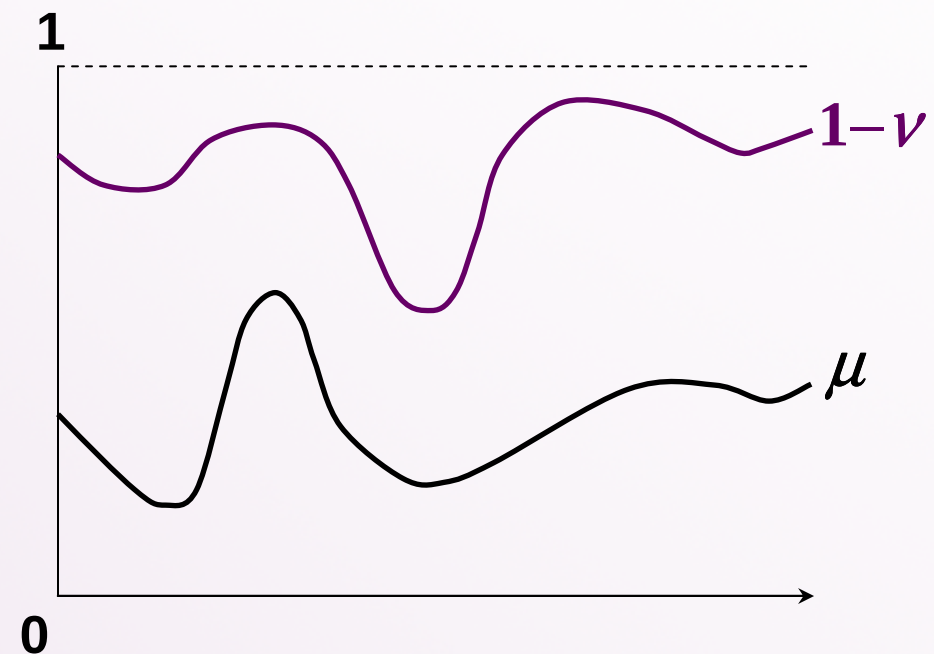
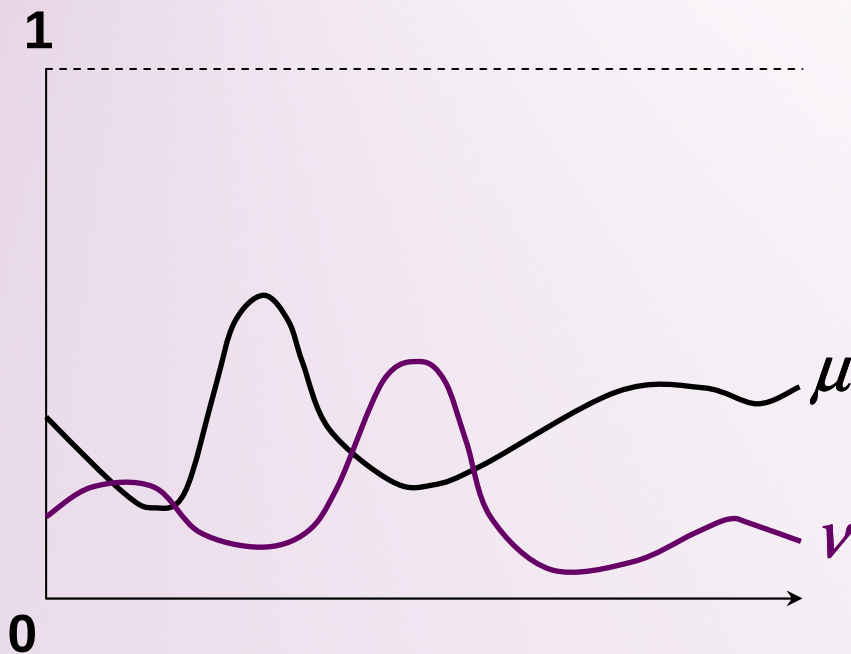
The value $\pi_A(x)$ is called the degree of non-determinacy (or degree of uncertainty) of the element $x \in X$ to the intuitionistic fuzzy set A .

Стойността $\pi_A(x)$ се нарича степен на неопределеност на елемента $x \in X$ към интуиционистки размитото множество A .

INTUITIONISTIC FUZZY SETS

ИНТУИЦИОНИСТКИ РАЗМИТИ МНОЖЕСТВА

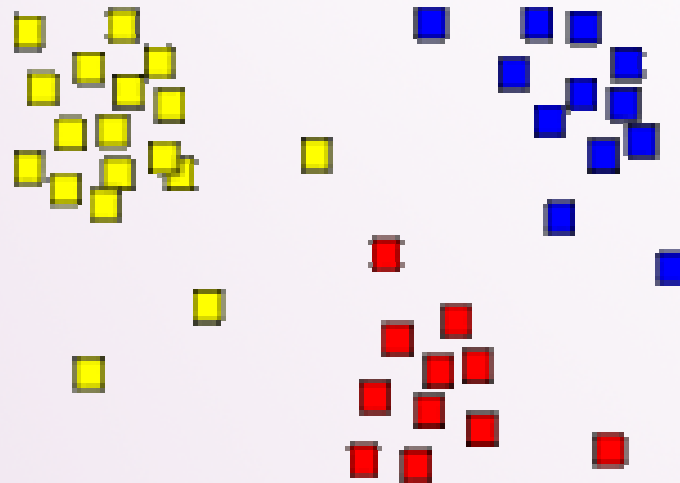
- Geometric interpretations of IFS
Геометрични интерпретации на ИРМ



CLUSTERING

КЛЪСТЕРИНГ

- A cluster is a small group or bunch of something.
- Клъстер е малка група или набор от неща.



CLUSTERING

КЛЪСТЕРИНГ

- Cluster analysis or clustering is the assignment of a set of observations into subsets (called clusters) so that observations in the same cluster are similar in some sense.
- Клъстерният анализ е разпределението на дадено множество от наблюдения в подмножества (клъстери), така че наблюденията от една извадка са сходни в някакъв смисъл.
- Data is divided into distinct clusters, where each data element belongs to exactly one cluster.
- Данните са разпределени в отделни клъстери, където всеки елемент принадлежи точно на един клъстер.

CLUSTERING

КЛЪСТЕРИНГ

- Distance measure - An important step in most clustering is to select a distance measure, which will determine how the similarity of two elements is calculated. This will influence the shape of the clusters, as some elements may be close to one another according to one distance and farther away according to another.
- Мярка за разстояние – Важна стъпка при клъстеринга обикновено е изборът на мярка за разстоянието, която определя как да се изчисли сходството на два елемента. Това е определящо за формата на клъстерите, тъй като някои елементи според една мярка могат да са по-близки, а според друга – по-далечни.

CLUSTERING

КЛЪСТЕРИНГ

Distance measures

- Euclidean distance

Разстояние по Евклид

- Manhattan distance

Разстояние по Манхатън

- Maximum norm

Максимална норма

- Mahalanobis distance

Разстояние по Махаланобис

- Hamming distance

Разстояние по Хеминг

CLUSTERING

КЛЪСТЕРИНГ

Types of clustering

- Hierarchical clustering

Йерархичен клъстеринг

- k-means clustering

Клъстерен анализ на k-средните

- Fuzzy c-means clustering

Размит клъстерен анализ на c-средните

- QT clustering algorithm

QT алгоритъм за клъстеринг

- Locality-sensitive hashing

Локално чувствително хеширане

- Graph-theoretic methods

Методи от теорията на графите

- Spectral clustering

Спектрален клъстеринг

CLUSTERING

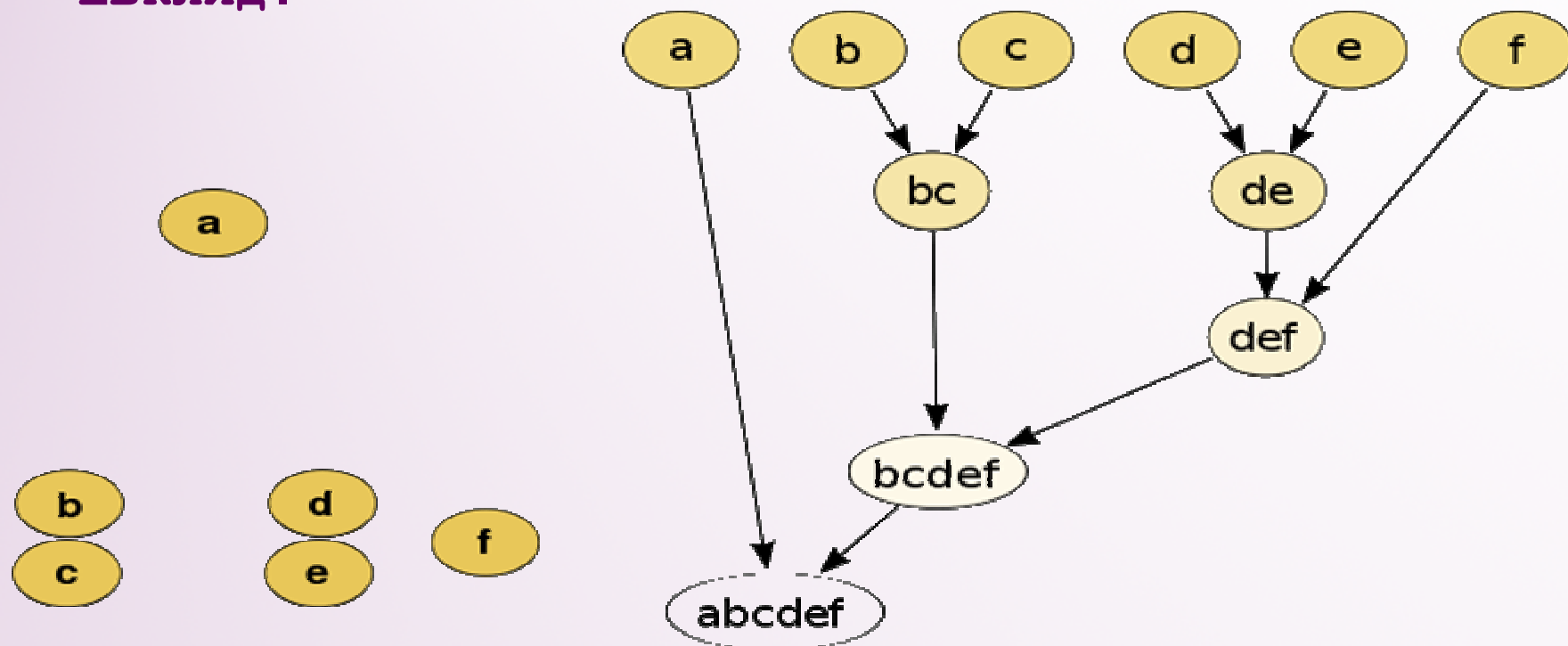
КЛЪСТЕРИНГ

- Hierarchical clustering creates a hierarchy of clusters which may be represented in a tree structure called a dendrogram.
- Йерархичният клъстеринг създава йерархия от клъстери, които могат да бъдат представени в дървовидна структура, наречена дендрограма.

CLUSTERING

КЛЪСТЕРИНГ

- Example. Hierarchical clustering with Euclidian distance.
- Пример. Йерархичен клъстеринг с разстояние по Евклид.



Merging clusters algorithm

- Step1: determine which elements to merge in a cluster
- Step2: construct a distance matrix - where the number in the i -th row j -th column is the distance between the i -th and j -th elements
- Step3: take the two closest elements, according to the chosen distance
- Step4: merge the clusters - The minimum distance between elements of each cluster -

$$\min\{d(x,y) \mid x \in \mathcal{A}, y \in \mathcal{B}\}$$

CLUSTERING

КЛЪСТЕРИНГ

- Suppose we have merged the two closest elements b and c , we now have the following clusters $\{a\}$, $\{b, c\}$, $\{d\}$, $\{e\}$ and $\{f\}$, and want to merge them further. Repeat the procedure.
- Да допуснем, че сме обединили двата най-близко разположени елемента b и c , и вече разполагаме със следните клъстери $\{a\}$, $\{b, c\}$, $\{d\}$, $\{e\}$ и $\{f\}$, и искаме да продължим да ги окрупняваме. Следователно, повтаряме процедурата.

CLUSTERING

КЛЪСТЕРИНГ

- The k-means clustering assigns each point to the cluster whose center (also called centroid) is nearest. The center is the average of all the points in the cluster – that is, its coordinates are the arithmetic mean for each dimension separately over all the points in the cluster.
- Клъстеринг на k-средните изпраща всяка точка към онзи клъстер, чийто център (центроид) е най-близо до нея. Центърът е точка, чиито координати са средно аритметичното по всяка координата на всички точки в клъстера.

CLUSTERING

КЛЪСТЕРИНГ

- k-means clustering algorithm.
- Step1: Choose the number of clusters, k .
- Step2: Randomly generate k clusters and determine the cluster centers, or directly generate k random points as cluster centers.
- Step3: Assign each point to the nearest cluster center, where "nearest" is defined with respect to one of the distance measures.
- Step4: Recompute the new cluster centers.
- Step5: Repeat the Steps 3&4 until some convergence criterion is met.

FUZZY CLUSTERING

РАЗМИТ КЛЪСТЕРИНГ

- In fuzzy clustering, each point has a degree of belonging to clusters, as in fuzzy logic, rather than belonging completely to just one cluster. Thus, points on the edge of a cluster, may be in the cluster to a lesser degree than points in the center of cluster.
- В размития клъстеринг, всяка точка има степен на принадлежност към отделните клъстерите, вместо на пълна принадлежност към един от тях. Така точките в периферията на клъстера му принадлежат в по-малка степен от точките в средата му.

FUZZY CLUSTERING

РАЗМИТ КЛЪСТЕРИНГ

- For each point x we have a coefficient giving the degree of being in the k -th cluster $u_k(x)$. Usually, the sum of those coefficients for any given x is defined to be 1.
- За всяка точка x имаме коефициент задаващ степента на принадлежност към k -тия клъстер $u_k(x)$. Обичайно сумата на тези коефициенти за всяка точка x се дефинира да бъде 1.

$$\forall x \left(\sum_{k=1}^{num.clusters} u_k(x) = 1 \right)$$

FUZZY CLUSTERING

РАЗМИТ КЛЪСТЕРИНГ

- With fuzzy c-means (FCM), the centroid of a cluster is the mean of all points, weighted by their degree of belonging to the cluster.
- При метода на размитите с-средни, центроидът на даден клъстер е средното на всички точки, претеглено спрямо степента на принадлежност на всяка точка към клъстера.

$$center_k = \frac{\sum_x u_k(x)^m x}{\sum_x u_k(x)^m}$$

FUZZY CLUSTERING

РАЗМИТ КЛЪСТЕРИНГ

- The degree of belonging is related to the inverse of the distance to the cluster center.
- Степента на принадлежност е реципрочна на разстоянието до центъра на клъстера.

$$u_k(x) = \frac{1}{d(\text{center}_k, x)}$$

FUZZY CLUSTERING

РАЗМИТ КЛЪСТЕРИНГ

- Then the coefficients are normalized and fuzzified with a real parameter $m > 1$ so that their sum is 1. So.
- После коефициентите се нормализират и размиват с реалния параметър $m > 1$, така че сумата им да стане равна на 1. Така

$$u_k(x) = \frac{1}{\sum_j \left(\frac{d(\text{center}_k, x)}{d(\text{center}_j, x)} \right)^{\frac{2}{m-1}}}$$

($m = 2$)

FUZZY CLUSTERING

РАЗМИТ КЛЪСТЕРИНГ

- **Algorithm**
- **Step1:** Choose the number of clusters.
- **Step2:** Assign randomly to each point coefficients for being in the clusters.
- **Step3:** Compute the centroid for each cluster using the formula above
- **Step4:** For each point, compute its coefficients of being in the clusters, using the formula
- **Step5:** Repeat the Steps 3&4 until some convergence criterion is met (that is, the coefficients' change between two iterations is no more than ε , the given sensitivity threshold)

INTUITIONISTIC FUZZY CLUSTERING

ИНТУИЦИОНИСТКИ РАЗМИТ КЛЪСТЕРИНГ

- A major challenge posed by real-world clustering applications is dealing with uncertainty in the localization of the feature vectors.
- Основно предизвикателство пред приложенията на клъстеринга в реалния свят е справянето с неопределеността при установяването на характеристикните вектори.

INTUITIONISTIC FUZZY CLUSTERING

ИНТУИЦИОНИСТКИ РАЗМИТ КЛЪСТЕРИНГ

- Considering that feature values may be subject to uncertainty due to imprecise measurements and noise, the distances that determine the membership of a feature vector to a cluster will also be subject to uncertainty.
- Предвид, че характеристикните стойности могат да са неопределени поради неточни измервания и информационен шум, разстоянията, които определят степента на принадлежност на даден характеристичен вектор към даден клъстер, също могат да са неопределени.

INTUITIONISTIC FUZZY CLUSTERING

ИНТУИЦИОНИСТКИ РАЗМИТ КЛЪСТЕРИНГ

- Therefore, the possibility of erroneous membership assignments in the clustering process is evident.
- Следователно е очевидна възможността за погрешни оценки на степените на принадлежност в рамките на процеса на клъстеринг.

INTUITIONISTIC FUZZY CLUSTERING

ИНТУИЦИОНИСТКИ РАЗМИТ КЛЪСТЕРИНГ

- The importance of the potential applications of intuitionistic fuzzy sets have drawn the attention of many researchers that have proposed various kinds of similarity measures between intuitionistic fuzzy sets.
- Значението на потенциалните приложения на интуиционистки размитите множества е привлякло вниманието на много учени, които са предложили различни видове мерки за сходство между интуиционистки размитите множества.

DISTRIBUTED CLUSTERING

РАЗПРЕДЕЛЕН КЛЪСТЕРИНГ

- Traditional Centralized Clustering methods require all data to be located at the place, where they are analyzed and cannot be applied in the case of multiple distributed datasets, unless all data are transferred to a single location and clustered.
- Традиционните методи за централизиран клъстеринг изискват всички данни да са локализирани на мястото, където ще бъдат анализирани, и са неприложими в случая с множество разпределени множества от данни, освен ако всички те не се пренесат на едно място, където да се извърши клъстерингът.

DISTRIBUTED CLUSTERING

РАЗПРЕДЕЛЕН КЛЪСТЕРИНГ

- Due to technical, economical or security reasons, it is not always possible to transmit all data from different local sites to single location and then perform global clustering.
- Поради съображения за сигурност, както и технически, икономически съображения, не винаги е възможно да се пренесат всички данни от различни източници на едно място, където да се извърши глобален клъстеринг.

DISTRIBUTED CLUSTERING

РАЗПРЕДЕЛЕН КЛЪСТЕРИНГ

- In the real world, distributed clustering applications frequently involve disparate datasets that also consist of inconsistencies or outliers, where it is difficult to obtain homogeneous and meaningful global clusters.
- В реалността, приложенията за разпределен клъстеринг често включват несъизмерими множества от данни, които съдържат несъгласуваности и външни елементи и при които е трудно да се получат хомогенни и смислово коректни глобални клъстери.

DISTRIBUTED CLUSTERING

РАЗПРЕДЕЛЕН КЛЪСТЕРИНГ

- It is obvious that alternate distributed clustering algorithms reduce the communication overhead, central storage requirements and computation times by exchanging few data and avoiding synchronization as much as possible.
- Очевидно е, че алтернативните алгоритми за разпределен клъстеринг намаляват ресурсите за комуникация, нуждата от централно хранилище на данните и времето за изчисление, като обменят малко данни и избягват синхронизацията доколкото могат.

APPLICATIONS OF CLUSTERING

ПРИЛОЖЕНИЯ НА КЛЪСТЕРИНГА

- **Biology**
Биология
- **Medicine, Psychology and Neuroscience**
Медицина, психология и невронауки
- **Market research**
Пазарни изследвания
- **Educational research**
Изследвания в образователната сфера
- **Social networks analysis**
Анализ на социални мрежи
- **Software evolution**
Еволюция на разработката на софтуер

APPLICATIONS OF CLUSTERING

ПРИЛОЖЕНИЯ НА КЛЪСТЕРИНГА

- Image segmentation
Сегментация на образи
- Data mining
Извличане на информация от данни
- Search results grouping
Пазарни изследвания
- Mathematical chemistry
Математическа химия
- Climatology, physical geography
Климатология, физическа география
- Crime analysis
Анализ на престъпления

*Thank you
for your attention!*



DR. R. PARVATHI

DEPT. OF MATHEMATICS, VELLALAR COLLEGE FOR WOMEN

ERODE, TAMILNADU, INDIA

paarvathis@rediffmail.com