



Algoritmi di Fuzzy Clustering

Andrea D'Amicis 869008

Introduzione

La seguente relazione mostra la modalità con cui sono stati individuati e in un certo senso classificati ai fini dei nostri interessi gli algoritmi di Fuzzy clustering.

Obiettivo e procedimento

Innanzitutto è stata utilizzata la piattaforma Scopus per estrarre tutte quelle pubblicazioni le quali avessero a che fare col fuzzy clustering attraverso un'interrogazione e quindi attraverso l'utilizzo della seguente query

```
TITLE ( fuzzy AND clustering AND ( survey OR review ) )
```

Una volta ottenuta la lista di articoli è stata fatta una analisi della descrizione (abstract) degli articoli ed è quindi stata fatta una scrematura. Le pubblicazioni ritenute come utili sono quelle che nominavano degli algoritmi di fuzzy clustering mentre tutte quelle pubblicazioni che parlavano in linea teorica del clustering o di domini di applicazione troppo specifici sono state scartate.

Successivamente sono, quindi, rimasti tutti quegli articoli che contenevano degli algoritmi che proponevano una soluzione su un dominio di dati generico e non

troppo specifico.

Una volta quindi ottenuto il pool di algoritmi, la fase successiva è stata quella di classificarli in modo da vedere se fossero correlati tra loro. Lo scopo era quindi capire se vi fosse un algoritmo “padre” da cui si diramassero altre versioni.

Scrematura iniziale

Di seguito vi è il risultato della query lanciata su Scopus. Gli articoli evidenziati in verde sono quelli che propongono un algoritmo riguardante il fuzzy clustering che sarà poi esaminato mentre gli altri non contenevano informazioni significative o parlavano in linea teorica del clustering. Quelli blu sono quelli in dubbio.

1. [A survey of fuzzy clustering validity evaluation methods](#)

Questo articolo, oltre a una breve descrizione del fuzzy clustering, parla dell'algoritmo FCVI e CFCVE senza dare dettagli. Cercando online ho trovato che:

- FCVI, algoritmo utilizzato per riconoscere i geni che mediano i differenti tipi di cancro (quindi questo lo escluderei)
- CFCVE, algoritmo trova il numero ottimale di cluster per il set di dati

2. [A Comprehensive Survey of Fuzzy Inference Systems Used for Clustering Problems in WSNs](#)

3. [Density-based IFCM along with its interval valued and probabilistic extensions, and a review of intuitionistic fuzzy clustering methods](#)

Questo articolo dice che la maggior parte degli algoritmi c-Means come FCM (Fuzzy c-Means), IFCM (Intuitionistic Fuzzy c-Means) e il PIFCM (Probabilistic Intuitionistic Fuzzy c-means) inizializzano in modo casuale i centroidi del cluster (prestazioni che dipendono molto dai centroidi iniziali). Propone come soluzione un algoritmo basato sulla densità: Density Based Probabilistic Intuitionistic Fuzzy c-Means (DPIFCM)

4. [Fuzzy C-Means clustering algorithm for data with unequal cluster sizes and contaminated with noise and outliers: Review and development](#)

Questo articolo parla invece dei problemi che si porta dietro FCM (fuzzy c means) riguardo ad outliers, rumore, valori anomali e alla disuguaglianza in termini di dimensioni tra i vari cluster che porta i centroidi dei cluster “piccoli” verso quelli “grandi”. Propone quindi una versione rivisitata in cui utilizza funzioni esponenziali adattive per eliminare gli impatti del rumore e dei valori anomali sui

centroidi dei cluster, modifica il vincolo dell'algoritmo FCM per impedire quindi ai cluster “grandi” o “più pesanti” di attrarre i centroidi di “piccoli” cluster verso di loro. Propone quindi come soluzioni i seguenti algoritmi:

- Possibilistic Fuzzy C-Means (PFCM)
 - Possibilistic C-Means (PCM)
 - Robust Fuzzy C-Means (FCM- σ)
 - Noise Clustering (NC), Kernel Fuzzy C-Means (KFCM)
 - Intuitionistic Fuzzy C-Means (IFCM)
 - Robust Kernel Fuzzy C-Mean (KFCM- σ)
 - Robust Intuitionistic Fuzzy C-Means (IFCM- σ)
 - Kernel Intuitionistic Fuzzy C-Means (KIFCM)
 - Robust Kernel Intuitionistic Fuzzy C-Means(KIFCM- σ)
 - Credibilistic Fuzzy C-Means (CFCM)
 - Size-insensitive integrity-based Fuzzy C-Means (siibFCM)
 - Size-insensitive Fuzzy C-Means (csiFCM)
 - Subtractive Clustering (SC)
 - Density Based Spatial Clustering of Applications with Noise (DBSCAN)
 - Gaussian Mixture Models (GMM)
 - Spectral clustering, and Outlier Removal Clustering (ORC)
5. Generating clustering-based interval fuzzy type-2 triangular and trapezoidal membership functions: A structured literature review
 6. A Survey on Kernelized Fuzzy Clustering Tools for Data Processing
 7. A Survey on Spherical Fuzzy Sets and Clustering the Literature
 8. Comparison of Fuzzy C-Means and K-Means Clustering Performance: An Application on Household Budget Survey Data
 9. Selection of Optimal Number of Clusters and Centroids for K-means and Fuzzy C-means clustering: A Review
 10. Different approaches for missing data handling in fuzzy clustering: A review

11. Survey on Fuzzy Associative Classifications Techniques and Their Performance Evaluation with Different Fuzzy Clustering Techniques Over Big Data
12. A unified framework of sentimental analysis for online product reviews using genetic fuzzy clustering with classification
13. A review and proposal of (fuzzy) clustering for nonlinearly separable data
14. Evolving fuzzy and neuro-fuzzy approaches in clustering, regression, identification, and classification: A Survey.
15. Image segmentation through fuzzy clustering: A survey.
16. Fuzzy C-means clustering strategies: A review of distance measures
17. Node clustering in wireless sensor networks using fuzzy logic: Survey.
18. Clustering for Energy Efficient and Redundancy Optimization in WSN using Fuzzy Logic and Genetic Methodologies a Review
19. Review on Clustering Algorithms Based on Data Type: Towards the Method for Data Combined of Numeric-Fuzzy Linguistics
20. A survey: Challenges of image segmentation based fuzzy c-means clustering algorithm
21. Clustering of nonstationary data streams: A survey of fuzzy partitional methods
22. A survey on image segmentation by Fuzzy C Means clustering techniques
23. A comprehensive review of fuzzy-based clustering techniques in wireless sensor networks
24. Fuzzy Clustering Algorithms - Review of the Applications
25. Survey on grid clustering approach with intuitionistic fuzzy histogram
26. Performance Analysis of Various Fuzzy Clustering Algorithms: A Review
Questo articolo elenca i seguenti algoritmi FCM, PCM, PFCM, FCM- σ , T2FCM, KT2FCM, IFCM, KIFCM, IFCM- σ , KIFCM- σ , NC, CFCM, DOFCM
27. A review on fuzzy logic based clustering algorithms for wireless sensor networks
28. Aspect based sentiment analysis for textual reviews using fuzzy C-means clustering algorithm
29. Fuzzy C-means (FCM) clustering algorithm: A decade review from 2000 to 2014
30. Survey on clustering methods: Towards fuzzy clustering for big data

31. A review on type-2 fuzzy logic applications in clustering, classification and pattern recognition
32. Erratum to: An efficient image clustering and content based image retrieval using Fuzzy K means clustering algorithm(International Review on Computers and Software, 9, 1,(147-153)).
33. Fuzzy-based clustering of web services' quality of service: A review
34. Survey of clustering algorithms using fuzzy logic in wireless sensor network
35. Fuzzy logic based clustering in wireless sensor networks: A survey
36. Fuzzy and crisp clustering methods based on the neighborhood concept: A comprehensive review
37. The fire surveys studies based on the fuzzy clustering and the neural network
38. Fuzzy clustering for microarray data analysis: A review
39. Image segmentation using fuzzy clustering: A survey.
Novel Fuzzy C-Means Clustering (NFCM), Fuzzy Local Information C-Means (FLICM) Clustering Algorithm and Improved Spatial Fuzzy C-Means Clustering (ISFCM) algorithm
40. A survey of fuzzy clustering algorithms for pattern recognition - Part II
41. A survey of fuzzy clustering algorithms for pattern recognition - Part I
42. A survey of fuzzy clustering
43. A review of fuzzy clustering methods

Algoritmi

- **FCM (Fuzzy C-Means)**

Fuzzy C-Means (FCM) è una tecnica di clustering dei dati in cui un set di dati è raggruppato in N cluster e ogni punto nel set di dati che appartiene a ogni cluster in una certa misura.

- **IFCM (Intuitionistic Fuzzy c-Means)**

IFCM è una tecnica di clustering che estende l'algoritmo Fuzzy C-Means introducendo il concetto di insiemi fuzzy intuizionistici. In IFCM, ogni punto dati è assegnato non solo a un cluster con un certo grado di appartenenza, ma anche a un cluster secondario con un grado di non appartenenza e ad anche un grado

di incertezza/ambiguità. A differenza di fuzzy C-Means lo scopo è di minimizzare proprio l'incertezza.

- **PIFCM (Probabilistic Intuitionistic Fuzzy c-means)**

PIFCM è un algoritmo di clustering che estende FCM e in cui ogni punto dati è assegnato a un cluster con un certo grado di appartenenza, determinato dalle misure probabilistiche e intuizionistiche dell'algoritmo. PIFCM assegna un valore di probabilità a ciascun oggetto per ogni cluster, quindi la somma dei valori di probabilità assegnati a ciascun oggetto per ogni cluster è pari a uno.

- **DPIFCM (Density Based Probabilistic Intuitionistic Fuzzy c-Means)**

DPIFCM è un algoritmo di clustering che combina il clustering basato sulla densità, clustering probabilistico e insiemi fuzzy intuizionistici. Ogni punto, infatti, può appartenere a più di un cluster, con il grado di appartenenza rappresentato come valore di probabilità. DPIFCM utilizza una funzione di densità dei dati per calcolare i valori di probabilità dei centroidi dei cluster stimando la densità per ogni regione dello spazio dei dati e utilizzando essa per calcolare i valori di probabilità dei centroidi dei cluster. Ha come vantaggio il fatto di gestire cluster non sferici (grazie alla densità).

- **PFCM (Possibilistic Fuzzy C-Means)**

PFCM (Possibilistic Fuzzy C-Means) è un algoritmo di clustering molto simile a FCM ma in cui il grado di appartenenza di un oggetto a un cluster viene calcolato utilizzando una funzione di appartenenza possibilistica invece di una funzione di appartenenza fuzzy. La funzione di appartenenza possibilistica rappresenta la probabilità che un oggetto appartenga a un cluster, sulla base di quanto bene l'oggetto si adatta al centroide del cluster.

- **FCM- σ (Robust Fuzzy C-Means)**

FCM- σ (Robust Fuzzy C-Means) è un algoritmo di clustering che è un'estensione dell'algoritmo Fuzzy C-Means. L'algoritmo FCM- σ è una versione robusta dell'algoritmo Fuzzy C-Means, che è meno sensibile ai valori anomali o al rumore nei dati. Raggiunge questo obiettivo utilizzando una funzione robusta, che riduce l'influenza dei valori anomali sui risultati del clustering. Utilizza un criterio di soglia basato sulla deviazione standard dei dati per identificare i dati anomali e li rimuove dal processo di clustering. In particolare, il parametro σ rappresenta il numero di deviazioni standard sopra cui un dato viene considerato anomalo. Inoltre, tramite una funzione di peso, assegna maggior peso ai dati

che sono meno sensibili agli effetti degli outlier. Questo aiuta a mitigare l'effetto degli outlier sui centroidi dei cluster e quindi a migliorare la qualità del clustering.

- **KFCM (Kernelized Fuzzy C-Means)**

KFCM (Kernel Fuzzy C-Means) è una variante dell'algoritmo Fuzzy C-Means (FCM) che utilizza il concetto di kernel per proiettare i dati in uno spazio di dimensione superiore, dove è possibile applicare l'algoritmo FCM in modo più efficiente e accurato. In particolare, l'algoritmo KFCM utilizza una funzione kernel per mappare i dati in uno spazio di dimensione superiore, dove la separabilità dei cluster può essere migliorata. La funzione kernel è una funzione matematica che calcola la similarità tra due punti nel dataset, in base alla loro distanza nello spazio originale. La funzione del kernel utilizzata in KFCM viene generalmente scelta in base alle caratteristiche dei dati analizzati.

- **KFCM- σ (Robust Kernelized Fuzzy C-Mean)**

KFCM- σ (Robust Kernel Fuzzy C-Mean) è una estensione degli algoritmi Kernel Fuzzy C-Means (KFCM) e FCM- σ (Robust Fuzzy C-Means). Quindi avremo che i dati vengono mappati in uno spazio di dimensione superiore tramite la funzione kernel e da lì vengono clusterizzati con FCM- σ dando meno peso agli outlier e ai valori anomali.

- **IFCM- σ (Robust Intuitionistic Fuzzy C-Means)**

IFCM- σ (Robust Intuitionistic Fuzzy C-Means) è un algoritmo di clustering che estende IFCM (Intuitionistic Fuzzy c-Means) ed FCM- σ (Robust Fuzzy C-Means) combinando quindi il concetto di insiemi fuzzy intuizionistici insieme alla robustezza. Ogni punto dati è assegnato quindi non solo a un cluster con un certo grado di appartenenza, ma anche a un cluster secondario con un grado di non appartenenza e anche un grado di incertezza/ambiguità con l'accortezza di avere una funzione di peso che consente di gestire al meglio i valori anomali/outliers.

- **KIFCM (Kernelized Intuitionistic Fuzzy C-Means)**

KIFCM (Kernel Intuitionistic Fuzzy C-Means) è un algoritmo di clustering che estende KFCM (Kernel Fuzzy c-Means) ed IFCM (Intuitionistic Fuzzy c-Means) combinando quindi il concetto di insiemi fuzzy intuizionistici insieme ai vantaggi di proiettare i dati in uno spazio di dimensione superiore. Ogni punto dati è assegnato quindi non solo a un cluster con un certo grado di appartenenza, ma anche a un cluster secondario con un grado di non appartenenza e anche un grado di incertezza/ambiguità con l'accortezza di mappare i dati in uno spazio

superiore e consentendo di applicare FCM ed ottenere una separabilità dei cluster migliore.

- **KIFCM- σ (Robust Kernelized Intuitionistic Fuzzy C-Means)**

KIFCM- σ (Robust C-Mean Intuitionistic Fuzzy Kernel Median) è un algoritmo di clustering che estende ed unisce i concetti di IFCM- σ (Robust Intuitionistic Fuzzy C-Means) e di KIFCM (Kernel Intuitionistic Fuzzy C-Means) avendo quindi il concetto di insiemi fuzzy intuizionistici insieme alla robustezza proiettando i dati in uno spazio di dimensione superiore per applicare FCM in modo più efficiente.

- **CFCM (Credibilistic Fuzzy C-Means)**

CFCM (Credibilistic Fuzzy C-Means) è un algoritmo di clustering che estende l'algoritmo Fuzzy C-Means (FCM) incorporando anche una nozione di credibilità. A differenza dell'FCM, in cui ogni punto appartiene a ciascun cluster con un grado di appartenenza, in CFCM ogni punto ha una credibilità rispetto a ciascun cluster.

- **siibFCM (Size-insensitive integrity-based Fuzzy C-Means)**

siibFCM (Size-insensitive integrity-based Fuzzy C-Means) è un algoritmo di clustering basato sulla teoria degli insiemi fuzzy che utilizza una strategia di penalizzazione per gestire la presenza di rumore e outlier nel dataset. L'algoritmo siibFCM utilizza la distanza euclidea per calcolare la somiglianza tra gli oggetti del dataset e la matrice di adiacenza fuzzy per rappresentare le relazioni tra gli oggetti e i cluster. Tuttavia, a differenza dell'algoritmo Fuzzy C-Means (FCM), siibFCM utilizza una strategia di penalizzazione basata sull'integrità fuzzy per ridurre l'effetto del rumore e degli outlier. In particolare, siibFCM considera l'integrità fuzzy degli oggetti e dei cluster, ovvero la coerenza tra la somiglianza di un oggetto con i cluster e la sua appartenenza ai cluster. Se un oggetto ha un'alta integrità fuzzy, la sua somiglianza con i cluster viene considerata più affidabile e la sua appartenenza ai cluster viene penalizzata di meno. Invece, se un oggetto ha un'alta appartenenza a cluster ma una bassa integrità fuzzy, viene maggiormente penalizzato.

- **csiFCM (Size-insensitive Fuzzy C-Means)**

csiFCM (Size-insensitive Fuzzy C-Means) è un algoritmo di clustering che estende FCM (Fuzzy C-Means) che utilizza una strategia di pesatura. Per gestire la problematica delle differenze di dimensione tra i cluster csiFCM introduce un parametro di pesatura che tiene conto della dimensione dei cluster, consentendo di creare cluster con dimensioni più equilibrate. In questo modo,

csiFCM è in grado di gestire dataset con oggetti di diverse dimensioni e varianze senza che gli oggetti più grandi abbiano un impatto maggiore sui centroidi dei cluster. L'algoritmo csiFCM utilizza la distanza euclidea per calcolare la somiglianza tra gli oggetti e i centroidi dei cluster e la matrice di adiacenza fuzzy per rappresentare le relazioni tra gli oggetti e i cluster.

- **T2FCM (Type-2 Fuzzy C-Means)**

T2FCM (Type-2 Fuzzy C-Means) è un algoritmo di clustering che estende l'algoritmo Fuzzy C-Means (FCM) utilizzando insiemi fuzzy di tipo 2 per gestire l'incertezza nell'appartenenza di ciascun punto a ciascun cluster. Essendo di tipo 2 la funzione di appartenenza di ciascun elemento dipende non solo dall'elemento stesso, ma anche dal contesto in cui si trova. Ciò consente di rappresentare in modo più preciso l'incertezza e l'ambiguità nei dati.

- **KT2FCM (Type-2 Fuzzy C-Means)**

KT2FCM (Kernelized Type-2 Fuzzy C-Means) è un algoritmo di clustering che estende T2FCM (Type-2 Fuzzy C-Means) che si basa sui concetti dei fuzzy set di tipo 2 e che proietta i dati in uno spazio di dimensione superiore per applicare FCM in modo più efficiente.

- **FLICM (Fuzzy Local Information C-Means)**

Fuzzy Local Information C-Means (FLICM) è un algoritmo di clustering che estende FCM (Fuzzy C-Means) utilizzando l'informazione locale per migliorare la qualità del clustering.

- **ISFCM (Improved Spatial Fuzzy C-Means Clustering)**

Improved Spatial Fuzzy C-Means (ISFCM) è un algoritmo di clustering che estende l'algoritmo Fuzzy C-Means (FCM) incorporando informazioni spaziali per migliorare la qualità del clustering. ISFCM utilizza una funzione di peso che attribuisce maggiore peso ai punti che sono vicini tra loro nello spazio.

- **CFCVE (Credibilistic Fuzzy Clustering by Validity Estimation)**

CFCVE (Credibilistic Fuzzy Clustering by Validity Estimation) è un algoritmo di clustering ibrido che combina i vantaggi dell'algoritmo Fuzzy C-Means (FCM) e dell'algoritmo di Validity Estimation (VE) per migliorare la qualità del clustering. L'obiettivo è quello di trovare il numero ottimale di cluster per i dati in input, utilizzando un criterio di validità del clustering basato sulle statistiche dei dati utilizzando un indice di validità basato sulla dispersione intra-cluster e sulla separazione inter-cluster per valutare la qualità del clustering.

Schema delle dipendenze

