INDICE

Capitolo 1

CENNI SULLA CLUSTER ANALYSIS

di Loredana Cerbara e Giampaolo Iacovacci

1.1 La classificazione: usi e scopi

I metodi di classificazione, detti metodi di ***cluster analysis*** o ***clustering***, hanno lo scopo di classificare le unità statistiche attraverso l'uso di procedure che, di solito, sono applicabili quando su ogni unità statistica sono state rilevate le modalità di M caratteri. Tali metodi si sono sviluppati fin dalla fine del XIX secolo e si valuta che gli algoritmi che sono stati elaborati fino ad oggi siano circa un migliaio. I motivi principali di tanto interesse per questo tipo di algoritmi sono essenzialmente due: 1) le tecniche di analisi dei gruppi sono largamente usate nei più svariati campi di ricerca (fisica, scienze sociali, economia, medicina, ecc.), in cui la classificazione dei dati disponibili è un momento essenziale nella ricerca di modelli interpretativi della realtà; 2) l'evoluzione degli strumenti di calcolo automatico ha consentito di affrontare senza difficoltà la complessità computazionale che è insita in molti dei metodi di classificazione e che in precedenza aveva spinto i ricercatori ad orientarsi verso quelle tecniche di analisi dei gruppi che erano più facilmente applicabili. Si è resa così possibile la produzione di diversi algoritmi di classificazione, sempre più complessi dal punto di vista computazionale, ma anche sempre più efficienti nel trarre informazioni dai dati attraverso una loro opportuna classificazione.

1.2. Definizioni

Gli autori non sono concordi nel definire un processo di clustering: secondo Sokal, consiste nel ripartire un insieme di unità elementari in modo che la suddivisione risultante goda di alcune proprietà considerate desiderabili; per altri studiosi classificare delle unità statistiche significa formare dei gruppi di unità in modo che le unità che sono assegnate allo stesso gruppo siano simili tra loro e che i gruppi siano il più possibile distinti tra loro (Gordon, 1988). Indipendentemente dalla definizione, in generale un metodo di classificazione è caratterizzato da due fattori:

a) una misura del grado di diversità tra le coppie di unità;
b) un algoritmo con cui procedere alla ricerca dei cluster.

Modificando uno o l'altro di questi fattori si possono produrre una gran quantità di metodi diversi dei quali sono state proposte diverse classificazioni alcune basate sul tipo di algoritmo adottato dal metodo, altre basate sul tipo di risultato da esso fornito.

La più diffusa è quella, basata sul tipo di algoritmo, che distingue tra metodi ***gerarchici*** e metodi ***non gerarchici.***

I primi sono metodi che producono raggruppamenti successivi ordinabili secondo livelli crescenti o decrescenti della distanza (o, viceversa, della similarità). Si tratta di procedura iterative che considerano tutti i livelli di distanza e i gruppi che si ottengono ad un certo livello di distanza sono contenuti nei gruppi ottenuti ad un livello di distanza inferiore. I metodi gerarchici si possono ulteriormente dividere distinguendo tra metodi ***agglomerativi*** e ***scissori.*** Sono ***agglomerative*** quelle tecniche che, partendo da n elementi distinti, producono di volta in volta un numero decrescente di clusters di ampiezza crescente, fino ad associare in un unico gruppo tutte le n unità di partenza. Viceversa, i metodi ***scissori*** ripartiscono gli stessi n elementi, inizialmente compresi in un unico insieme, in gruppi sempre più piccoli e numerosi, finché il numero di clusters viene a coincidere con il numero delle unità. Tra i due approcci, quello agglomerativo è stato sicuramente privilegiato: queste tecniche sono infatti più semplici da programmare e, come è stato osservato, comportano un minor rischio di pervenire a suddivisioni delle unità che non rispecchiano l'effettiva struttura dei dati, al contrario, i metodi scissori possono più facilmente realizzare allocazioni sbagliate delle unità, che però non vengono corrette se in seguito non sono previste particolari procedure di aggiustamento.

Quando l'algoritmo produce un'unica suddivisione dell'insieme di partenza, considerata ottimale rispetto al criterio adottato, la classificazione risultante è ***non gerarchica.*** Appartengono a questa categoria tutte le classificazioni prodotte da un metodo di programmazione matematica o quelle che, tentando di migliorare una suddivisione provvisoria delle unità, effettuano una serie di riallocazioni finché non risulta soddisfatto un dato criterio di ottimalità.

I metodi non gerarchici dipendono in generale da due fattori:

a) presenza o assenza di centri;
b) esistenza o meno di una funzione obiettivo.

Queste suddivisioni in realtà non comprendono tutti i vari tipi di metodi, ma riescono comunque a classificare quelli più usati.

Esistono poi infinite versioni di uno stesso metodo, quando, pur applicando una stessa procedura di clustering, vengono utilizzate differenti distanze.

Ciò spiega come mai i pochi metodi proposti in principio si siano moltiplicati fino a costituire un campo molto vasto e complesso, a cui i diversi schemi logici di unificazione hanno tentato di dare un ordine.

Senza voler approfondire tale approccio, estremamente rigoroso ma scomodo a causa della complicata terminologia di cui si avvale, possiamo più semplicemente fare una prima distinzione tra gli algoritmi *esatti* e quelli *euristici*. Sono esatte quelle procedure che determinano una suddivisione delle n unità in c clusters, la quale risulta ottima rispetto alla misura di omogeneità dei gruppi, o a quella di similarità delle unità, ossia genera la migliore tra tutte le possibili partizioni di n elementi in c clusters. Gli algoritmi euristici, o non esatti, danno luogo ad una suddivisione buona o approssimativamente ottima, ma che tuttavia si discosterà in qualche misura dall'essere la migliore possibile.

Si comprende come le tecniche più diffuse appartengano a questa seconda categoria: esse sono infatti computazionalmente più efficienti di quelle esatte, le quali per esaminare tutte le possibili partizioni necessitano spesso di un numero di operazioni elementari che cresce in maniera esponenziale con n.

Oltre alle differenziazioni basate sul tipo di algoritmo, i vari metodi si possono distinguere anche in base alla classificazione che essi producono. I risultati di una classificazione si possono rappresentare attraverso una matrice con tante righe quante sono le unità e tante colonne quanti sono i gruppi: se abbiamo n unità e G gruppi la matrice è di dimensione $(n \times G)$ e contiene i valori di una ***funzione di appartenenza***. Tale funzione, indicata con $\boldsymbol{\mu_{ig}}$, è una funzione a G valori (dove G è il numero di gruppi della partizione o del ricoprimento) che associa ad ogni unità G numeri ognuno dei quali esprime il grado di appartenenza dell'unità *i*-esima al *g*-esimo gruppo (con i=1,2,...,n e g=1,2,...,G). L'intervallo di definizione di tale funzione permette di distinguere tra metodi di classici e metodi non classici detti anche *sfocati*: per i metodi classici la funzione di appartenenza è definita nell'insieme {0,1}, cioè assume solo i due valori 1 e 0, che indicano, rispettivamente, se una unità appartiene o non all'insieme; per i metodi sfocati l'insieme di definizione è l'intervallo [0,1] e quindi la funzione di appartenenza esprime il ***grado*** con cui una unità appartiene ad un gruppo. Se si distinguono inoltre i raggruppamenti, ottenuti in base ai vari metodi, in ***partizioni***, che sono i raggruppamenti con la caratteristica di rispettare il vincolo

$$S = \sum_{g=1}^{G} m_{ig} = 1 \qquad (1)$$

e ***ricoprimenti***, che sono i raggruppamenti con la caratteristica di rispettare il vincolo

$$S = \sum_{g=1}^{G} m_{ig} \geq 1 \qquad (2)$$

si ottiene la suddivisione seguente (Ricolfi, 1992):

Tabella 1.1: ***Distinzioni dei vari metodi basate sui risultati da essi forniti.***

		Metodi classici	Metodi sfocati
	Grado di appartenenza	{0,1}	[0,1]
Partizioni	$\sum_{g=1}^{G} m_{ig} = 1$	Classificazione classica	Classificazione sfocata
Ricoprimenti	$\sum_{g=1}^{G} m_{ig} \geq 1$	Classificazione sovrapposta	Classificazione sovrapposta sfocata

Da questo schema si evince che con metodi di ***classificazione classica***, si intendono tutti quei metodi che forniscono una partizione classica (cioè una suddivisione delle unità in gruppi tra loro disgiunti e tali che la loro unione fornisca l'insieme di tutte le unità). Con i metodi di ***classificazione sovrapposta*** si indicano i metodi che forniscono una suddivisione delle unità in gruppi non disgiunti, cioè tali che una medesima unità possa appartenere a più di un gruppo (ricoprimento classi-

co dell'insieme delle unità). Con metodi di ***classificazione sfocata*** indicheremo quei metodi che suddividono l'insieme delle unità in modo che una unità può appartenere solo in parte ad un gruppo e quindi per la parte rimanente appartiene ad altri gruppi. Infine si può pensare ad un'ultima categoria di metodi, che risulta dall'unione di queste due ultime, la quale fornisca dei ***ricoprimenti sfocati***, cioè dei gruppi sfocati sovrapposti. A questa categoria potremmo dare nome di metodi di ***classificazione sovrapposta sfocata***. Naturalmente, all'interno di queste distinzioni, valgono ancora quelle fatte in precedenza tra metodi gerarchici e non gerarchici per cui, ad esempio, esistono tecniche gerarchiche e tecniche non gerarchiche di classificazione classica, sovrapposta, sfocata e sovrapposta sfocata.

Le tecniche di classificazione di solito utilizzate sono quelle che forniscono partizioni classiche dell'insieme iniziale, mentre per gli altri tre tipi il numero di algoritmi a disposizione è piuttosto ridotto.

Ci sono inoltre altri metodi (Ponsard, 1985, Fustier, 1980) che danno una iniziale sfocatura ai dati assegnando all'inizio del procedimento una funzione ad ogni unità. Tale funzione, che chiameremo ***funzione caratteristica***, misura la quantità di carattere posseduta da una unità rispetto a quella posseduta dalle altre unità. Perciò tali metodi non usano i dati di partenza per classificare le unità, ma li sostituiscono con dei dati 'sfocati'. In tal modo si possono utilizzare procedure di classificazione, che sono applicabili solo a caratteri misurabili, anche se si dispone di dati di qualunque natura. Le classificazioni che ne derivano sono però delle classificazioni classiche o al più sovrapposte, in quanto, secondo le definizioni appena date, la funzione di appartenenza finale è a valori in {0,1}. Non bisogna quindi confondere la classificazione sfocata con le classificazioni che si ottengono a partire da dati sfocati ma che presentano una funzione di appartenenza a valori in {0,1} e pertanto sono metodi da assegnare alla categoria di quelli classici.

In questa trattazione non ci soffermeremo sulla teoria degli insiemi sfocati, perché sono molti gli autori che si sono occupati della logica sfocata in modo rigoroso: pensiamo a Zadeh, uno tra i promotori di questa logica, a Ruspini, a Kaufmann, a Leung, a Ponsard e Tran qui solo per citarne alcuni tra i più autorevoli. Perciò faremo riferimento solo ai concetti di logica sfocata, che di volta in volta, sarà più opportuno chiarire e rimandiamo a questi autori per ulteriori approfondimenti.

1.3. La classificazione classica

In questo gruppo abbiamo compreso tutti quei metodi di analisi dei gruppi che producono una partizione classica delle unità, cioè suddividono le unità in gruppi disgiunti e tali che la loro unione fornisca l'insieme di tutte le unità. In questo paragrafo ci limitiamo a dare qualche accenno solo sui metodi più usati.

Tra i metodi gerarchici, quelli più usati sono quelli di tipo agglomerativo. Ad esempio:

a) il ***metodo del legame singolo (SLM)***, che si basa sulle distanze tra le unità: le unità che sono le une rispetto alle altre a distanza minima vengono assegnate ad un unico gruppo; si calcola poi la distanza tra questo gruppo appena formato e le rimanenti unità come la *minima distanza* tra le unità del gruppo e le altre unità. Se si sono già formati dei gruppi si calcola la distanza tra il gruppo appe-

na formato e gli altri gruppi come la minima distanza tra le unità del gruppo appena formato e le unità degli altri gruppi. Si ripete il procedimento fino a che tutte le unità sono nello stesso gruppo.

b) il ***metodo del legame completo (CLM)***, che si basa su un algoritmo del tutto simile a quello del legame singolo con la sola differenza che la distanza tra il gruppo appena formato e ognuna delle rimanenti unità (o gruppi) è calcolata come la *massima distanza* tra le unità del gruppo e le rimanenti unità.

c) il ***metodo del legame medio (ALM)***, che procede come i precedenti, ma calcola la distanza tra un gruppo ed una unità come la distanza tra l'unità e una unità fittizia in cui ciascun carattere è presente con una media delle modalità presentate dalle unità comprese nel gruppo.

E ancora: metodo del centroide, metodo della mediana, metodo della varianza minima, metodo del legame flessibile e tanti altri, su cui non ci soffermiamo.

Tra i metodi non gerarchici, i più usati sono quelli detti ***metodi di suddivisione iterativa*** e i ***metodi di programmazione matematica***. In genere, questi metodi partono da una iniziale suddivisione delle unità e procedono spostando le unità da un gruppo all'altro fino a che non si raggiunge una situazione ottimale che non consente altri spostamenti.

I metodi di suddivisione iterativa eseguono degli spostamenti effettivi delle unità: si calcolano i centroidi dei vari gruppi (oppure si scelgono dei nuclei o dei semi intorno ai quali si devono raggruppare le unità) e si assegna ogni unità al gruppo più vicino; poi si ricalcolano di nuovo i centroidi e si ripete il procedimento fino a che non si possono più spostare le unità. In alcuni casi i metodi sono dotati anche di una funzione obiettivo che valuta la bontà di una determinata partizione in modo che si possa scegliere lo spostamento più conveniente tra quelli possibili, come nel caso del ***metodo delle k-medie (HCM).***

I metodi di programmazione matematica si basano invece su spostamenti virtuali delle unità, fatti secondo la soluzione di un problema minimo o massimo vincolato, e non contemplano il calcolo dei centroidi dei gruppi.

Il punto debole di questo tipo di metodi non gerarchici sta nelle scelte che devono essere compiute all'inizio: si deve fare una iniziale partizione delle unità, si devono scegliere i nuclei o semi e, inoltre, si deve scegliere la funzione obiettivo. E' chiaro che scelte iniziali diverse porteranno inevitabilmente a diverse partizioni finali e che questi procedimenti possono diventare inaffidabili a meno che non si tenti di superare l'inconveniente di dover fare delle scelte iniziali (L. Ricolfi, 1992).

Ma in generale si pone il problema della scelta del metodo da adottare, dal momento che i diversi metodi di classificazione portano in genere a soluzioni diverse. Perciò è stato proposto da diversi autori di costruire un'unica classificazione aggregando le classificazioni che possono risultare dall'applicazione di diverse procedure ad uno stesso insieme di dati. Si tratta delle tecniche dette del ***consenso*** che producono una classificazione unica a partire da più classificazioni. Si tratta di tecniche che ultimamente hanno avuto nuovi sviluppi perché sono state applicate allo studio delle matrici di dati a *tre vie* (cioè formate da unità, variabili e occasioni) in quanto sono ottimi strumenti di sintesi dell'informazione (Gordon e Vichi, 1997).

1.4. La classificazione sovrapposta

Questo tipo di analisi dei dati, nato nel contesto della teoria dei grafi, rappresenta il tentativo di superare i problemi che nascono quando si presenta il caso di unità con caratteristiche intermedie a due o più gruppi (caso, in realtà, piuttosto frequente) e sarebbe opportuno assegnare l'unità ad entrambi i gruppi. In tal modo non si ottiene una partizione delle unità, ma un ricoprimento e i gruppi formati vengono chiamati di solito ***clump*** (Gordon, 1981) anziché cluster, e la tecnica di analisi dei dati viene detta clumping o classificazione sovrapposta (overlapping clustering).

Ai metodi che generano classificazioni sovrapposte è stata dedicata un'attenzione molto minore rispetto ai metodi di classificazione classica anche se essi sono di notevole interesse teorico. Ciò è dovuto in parte al fatto che molti autori fanno una certa confusione tra l'approccio sfocato e l'approccio sovrapposto, per cui alcuni algoritmi che si definiscono come algoritmi di classificazione sfocata, in realtà sono solo di classificazione sovrapposta. D'altra parte alcuni metodi di clumping finiscono per sovrapporre i gruppi in modo eccessivo e il ricoprimento che si ottiene diventa impossibile da interpretare.

Ma anche per questo problema sono state proposte delle soluzioni, come l'introduzione di vincoli per limitare le sovrapposizioni consentite imponendo, ad esempio, agli insiemi un numero massimo di sovrapposizioni pari a (k-1).

Vale la pena di ricordare il ***metodo delle piramidi*** (Diday, 1986), che è un metodo gerarchico di classificazione sovrapposta. Scelto un indice di similarità tra clump (che è anche un indice di similarità tra unità in quanto una unità è vista come un clump formato da un solo oggetto), si aggregano ad ogni stadio i clump più vicini formando proprio una struttura a piramide che ha alla base le unità ancora non raggruppate e il cui vertice rappresenta il clump finale, cioè quello che comprende tutte le unità. Anche per questo metodo sono state proposte delle limitazioni che servono ad ottenere un numero ridotto di sovrapposizioni: l'unione di due gruppi non è sempre consentita (se i e j sono gli elementi estremi di un gruppo g, cioè gli elementi che sono alla base del clump e delimitano il clump stesso, non si può unire a g nessun gruppo che non contenga sia i che j).

1.5. La classificazione sfocata

L'interesse teorico di questi metodi è dovuto al fatto che essi trattano bene l'imprecisione: le unità statistiche non sono sempre classificabili con esattezza perché non è raro il caso di unità che possono essere assegnate indifferentemente a più gruppi. Questi metodi assegnano ogni singola unità in parte a ciascun gruppo in modo che la classificazione che risulta non solo mostri come aggregano le unità, ma riesca anche a mostrare quanto una unità appartiene ad un gruppo. In tal modo l'assegnazione di una unità ad un gruppo non è mai una forzatura mentre la non assegnazione di una unità ad un gruppo indica con certezza che quella unità non appartiene a quel gruppo.

I metodi di classificazione sfocata non hanno quindi la pretesa di dare risposte precise su come si aggregano i dati, cosa che si può fare più agevolmente con un metodo di analisi classica, ma, al contrario, tentano di rappresentare proprio

l'imprecisione insita nei dati. Perché, come dice Ponsard (1985), probabilmente solo un modello *impreciso* della realtà è in grado di rappresentarla più esattamente di quanto possa fare un modello *preciso*.

I metodi di classificazione sfocata, come abbiamo detto, producono delle partizioni o dei ricoprimenti dell'insieme dei dati, ma la funzione di appartenenza di ogni unità ai vari gruppi non assume, come nei metodi di classificazione classica, solo i valori 0 e 1, ma assume un valore compreso tra 0 e 1 e misura il grado di appartenenza dell'unità ad un gruppo. La partizione (o ricoprimento) che si ottiene si chiama partizione sfocata (o ricoprimento sfocato).

Anche questi metodi possono essere distinti in ***gerarchici,*** che sono descritti in dettaglio nel capitolo 2, e ***non gerarchici*** che sono descritti nel terzo capitolo.

Saranno inoltre forniti degli esempi concreti di applicazione dei metodi di classificazione sfocata a problematiche di tipo socio-economico, in modo che questo volume possa rappresentare uno strumento di lavoro, un piccolo manuale d'uso, per i ricercatori e gli studiosi che intendono servirsi di tecniche di classificazione sfocata.

Capitolo 2

METODI GERARCHICI DI CLASSIFICAZIONE SFOCATA

di Loredana Cerbara

2.1. Introduzione

I metodi di classificazione sfocata che descriveremo in questo capitolo si caratterizzano per il fatto che si tratta di procedure gerarchiche che si compongono di due fasi: nella prima si calcola una misura della similarità tra le coppie di unità mentre nella seconda si applica alla matrice delle similarità ottenuta una procedura di classificazione delle unità che si conclude con l'assegnazione, a ciascuna unità, di una funzione di appartenenza ai gruppi che si sono formati. Naturalmente si potrebbero escogitare un gran numero di metodi di classificazione sfocata, analogamente a quanto è accaduto per i metodi di classificazione classica. Perciò i metodi descritti di seguito possono essere modificati per produrre tutte le possibili varianti di essi che meglio si adattano a situazioni particolari.

Di ogni metodo descritto verrà anche fornita un'applicazione d'esempio e, quando possibile, sarà dato spazio anche all'analisi dei risultati in modo che risulti facilitata la comprensione dei metodi e delle potenzialità di applicazione alle problematiche affrontate nei diversi ambiti della ricerca.

2.2.1. Il metodo della sintesi di più partizioni (Zani, 1989)

E' un procedimento che, partendo da M partizioni iniziali delle unità, arriva ad una classificazione sfocata di esse.

Supponiamo di avere n unità spaziali su cui siano rilevati M caratteri di tipo misto (quantitativi e/o qualitativi). Per determinare la similarità tra una coppia di unità si fa anzitutto, per tutti i caratteri considerati, una iniziale partizione in gruppi delle unità e si assume come indice di similarità tra le due unità che costituiscono la coppia, la frequenza relativa delle partizioni, una per ogni carattere, in cui le due unità si trovano incluse in uno stesso gruppo. Agli indici di similarità così ottenuti si applica poi una procedura di classificazione, molto simile alle procedure classiche, che produce una successione gerarchica di partizioni, cioè, per ogni livello di similarità considerato, la procedura produce una partizione sfocata delle unità i cui gruppi contengono quelli formati ai livelli precedenti di similarità. Nel corso di tale procedura ad ogni livello di similarità si assegna a ciascuna unità la sua funzione di appartenenza alla partizione sotto il vincolo che la somma dei gradi di appartenenza di ogni unità a tutti i gruppi sia uguale a 1 (vincolo necessario per ottenere una partizione sfocata anziché un ricoprimento, Bezdek, 1981).

Si ha quindi il problema della scelta della partizione iniziale di ciascun carattere. Tale scelta è però abbastanza naturale per i caratteri sconnessi o ordinali o per quelli quantitativi discreti con un numero ridotto di modalità in quanto sono le stesse modalità che comportano generalmente per ogni carattere la partizione migliore. Per i caratteri quantitativi le cui modalità non possono che essere classi (o perché il carattere è continuo o perché è discreto ma con un numero molto alto di modalità), la partizione in classi presenta indubbiamente un più ampio margine di soggettività. E' quindi almeno consigliabile (secondo S. Zani) ripetere la procedura considerando più partizioni iniziali con riferimento allo stesso carattere confrontando di conseguenza più risultati tra loro.

Per i caratteri quantitativi esistono comunque diversi criteri per la ricerca della partizione iniziale; possiamo, ad esempio, citarne alcuni:

i) suddivisione in base ai quartili;
ii) metodo della minima varianza (Spath, 1985);
iii) metodo delle classi naturali di Mineo (1978);
iv) metodo di ottimizzazione di Butler (1988).

Se si adotta il primo di questi criteri, cioè quello della suddivisione in quartili, le unità vengono raggruppate assegnando al primo gruppo quelle per le quali la modalità del carattere è compresa tra il minimo e il primo quartile, al secondo gruppo le unità con modalità del carattere compresa tra il primo e il secondo quartile, e così via.

Si individuano così le M partizioni indotte dalle M variabili.

Siano g_k i gruppi della partizione indotta dal k-esimo carattere (con $2 \leq g_k \leq n-1$ e $k = 1,2,...,M$).

Siamo in grado di definire il ***grado di appartenenza congiunto*** (ovvero la similarità) di due unità statistiche, a_i e a_j: esso è dato dalla frequenza relativa delle partizioni in cui le due unità sono incluse in uno stesso gruppo:

$$z_{ij} = \frac{1}{M} \sum_{k=1}^{M} d_k(i,j) \qquad (2.1)$$

dove $d_k(i,j)$ vale 1 se a_i e a_j sono nello stesso gruppo nella partizione k-esima, e vale 0 altrimenti.

z_{ij} è un indice di similarità che assume valori nell'intervallo [0,1], è simmetrico (perché $z_{ij}=z_{ji}$), gode della proprietà riflessiva (perché $z_{ii}=z_{jj}$) ed il suo massimo corrisponde alla piena appartenenza delle due unità allo stesso gruppo in ciascuna delle M partizioni ($z_{ij} \leq z_{ii}=1$).

Con i valori dell'indice z_{ij} si costruisce la **matrice delle similarità**. Per ottenere le classificazioni sfocate, si possono applicare alla matrice delle similarità diversi metodi di cluster per i quali occorre però definire il grado di appartenenza di una unità ad un gruppo sfocato. Esso può essere:

a) il massimo delle similarità tra essa e ciascuna delle altre unità incluse nel gruppo;
b) il minimo delle similarità tra essa e ciascuna delle altre unità incluse nel gruppo.

Mentre la definizione b) presenta analogie con il metodo di classificazione classica del legame completo, la definizione a) presenta analogie con il metodo di clas-

sificazione classica del legame singolo, in quanto una unità viene inclusa in un gruppo sfocato, ad un certo livello, quando essa presenta un valore della similarità con almeno un elemento del gruppo pari al valore del livello.

La procedura di classificazione si articola in tre fasi:

i) Si riuniscono tra loro le eventuali coppie di unità con similarità uguale ad 1, ottenendo una partizione delle unità. Si dice, in tal caso, che queste due unità costituiscono un "*nucleo*" (core) (Rolland-May, 1985). I nuclei possono essere ovviamente più di uno.

ii) Nella matrice delle similarità si prende in considerazione la similarità α uguale a (M-1)/M e si individuano le coppie che presentano tale grado di similarità. Se entrambe le unità costituenti una coppia con similarità α non sono già state assegnate ad un gruppo precedente, esse vengono a formare un nuovo gruppo, con grado di appartenenza uguale ad α (tali unità individuano un "*nucleo al livello a*"); se una delle unità era già stata inserita in un gruppo, l'altra viene aggregata a quel gruppo, con grado di appartenenza al gruppo medesimo uguale ad α.

iii) Si itera la fase sub ii) considerando livelli via via decrescenti del grado di appartenenza: (M-2)/M, (M-3)/M, ... In questi passi successivi si può manifestare anche il caso di coppie di unità che presentano tra loro il grado di appartenenza, considerato a quel livello, ma che sono già state inserite in gruppi diversi. In questa circostanza S. Zani suggerisce di assegnare "in parte" ciascuna unità all'altro gruppo, con grado di appartenenza uguale al valore di α in oggetto, purché la somma per riga dei gradi di appartenenza risulti uguale a 1; in caso contrario si potrà attribuire convenzionalmente come grado di appartenenza il valore massimo che soddisfa tale vincolo. Se tale valore massimo è 0 l'unità non può più essere assegnata a nessun gruppo anche se presenta il necessario livello di similarità con altre unità.

L'algoritmo genera dunque una successione gerarchica di raggruppamenti sfocati, in corrispondenza di livelli decrescenti della similarità (Dimitrescu,1988).

2.2.2. Un'applicazione

Una semplice applicazione di questo metodo è fornita dallo stesso Zani (1988). Vogliamo però riportare una breve applicazione (purtroppo di questa tecnica non sono state fatte applicazioni che non fossero soltanto d'esempio) di M. A. Milioli (1994) di tipo territoriale con dati demografici, che forse risulta di maggiore interesse per i lettori di questo volume.

Sono state classificate con il *metodo della sintesi di più partizioni* le regioni italiane in base ai valori assunti da 6 indicatori demografici riportati di seguito:

Tabella 2.1. ***Indicatori demografici.***	
X_1	Incidenza della popolazione infantile
X_2	Incidenza della popolazione senile
X_3	Indice di vecchiaia
X_4	Indice di dipendenza dei giovani
X_5	Indice di fecondità
X_6	Indice di mortalità
Fonte: Istat, 1990	

Sono state considerate le partizioni sfocate ottenute a diversi livelli di similarità, ma alla fine è stato scelto il livello a=0.5 perché i gruppi mantenevano ancora la loro specificità e le regioni che rimanevano isolate erano in numero trascurabile. La partizione sfocata ottenuta è riportata nella tabella 2.2 in cui tra parentesi è indicato il grado di appartenenza. Le regioni per le quali la somma dei gradi di appartenenza risulta minore di 1 per la parte rimanente costituiscono unità isolate, così come sono isolate quelle regioni che non sono presenti affatto nella tabella 2.2.

Tabella 2.2. ***Classificazione ottenuta al livello di similarità 0.50.***	
Gruppi	**Regioni**
1	Val d'Aosta (1), Lombardia (0.83), Umbria (0.83), Marche (1)
2	Trentino A. A. (1), Lazio (1), Abruzzi (0.83), Molise (0.83), Basilicata (0.33), Sicilia (0.17), Sardegna (0.5)
3	Piemonte (0.83), Friuli (1), Liguria (1), Emilia Romagna (1), Toscana (1)
4	Campania (1), Puglia (1), Basilicata (0.67), Calabria (1), Sicilia (0.83), Sardegna (0.5)
Fonte: M. A. Milioli, 1994	

Come si può vedere, si tratta di una partizione di semplice lettura: si sono formati soltanto quattro gruppi e le sovrapposizioni sono relativamente poche. C'è in generale una differenziazione tra nord e sud in quanto i gruppi 1 e 3 sono formati esclusivamente da regioni del centro-nord, mentre il gruppo 4 è formato solo da regioni del sud. Il gruppo 2, invece, mostra una sovrapposizione tra queste due zone a dimostrazione del fatto che, nonostante la diversità evidente del sud rispetto al resto del paese, ci sono delle caratteristiche che si ritrovano in tutto il paese, isole comprese, anche se si nota che i gradi di appartenenza vanno abbassandosi man mano che si procede da nord a sud, come ad evidenziare che si tratta di caratteristiche più decisamente presenti al nord rispetto al sud.

Trattandosi di un piccolo esempio non siamo in grado (e forse non è neanche interessante) di eseguire un'analisi dei risultati più approfondita. Gli esempi dei paragrafi seguenti, invece, sono tratti da applicazioni demografiche di maggior interesse e saranno commentati con maggior ampiezza.

2.3.1. Il metodo dei ricoprimenti sfocati

Si tratta di un metodo gerarchico di classificazione sfocata basato sulla costruzione, e successiva scomposizione, di una matrice delle similarità che fornisce ricoprimenti sfocati delle unità di partenza.

Occorre dunque definire l'indice di similarità adottato in questo metodo. Tale indice si adatta al minimo livello di misura dei caratteri che si hanno a disposizione per cui non costringe né a considerare soltanto variabili quantitative, come spesso accade quando si utilizzano i metodi di cluster analysis, né ad effettuare manipolazioni dei dati che comportano generalmente variazioni degli indici di similarità. Supponiamo di avere un collettivo di n unità su cui si rilevano m caratteri:

a) se tutti i caratteri sono quantitativi, per ogni carattere si calcola la distanza relativa (cioè la distanza rapportata alla massima distanza assunta per quel carattere) tra le coppie di unità e se ne fa il complemento ad uno; si ha cioè, per il generico carattere k-esimo (con k=1,2,....,m):

$$V_{ij}(k) = 1 - \frac{d_{ij}(k)}{\max_{i,j}\left[d_{ij}(k)\right]} \qquad (3.1)$$

dove $d_{ij}(k)$ è un indice di distanza detto *distanza di Hamming generalizzata* calcolato tra le unità i-esima e j-esima relativamente al carattere k-esimo; l'indice di similarità complessiva tra le unità i e j risulta essere:

$$S(i,j) = \sum_{k=1}^{m} V_{ij}(k) \cdot p(k) \qquad (3.2)$$

dove p(k) è il peso non negativo che, nel caso in cui si possa ritenere di poter dare pesi diversi ai caratteri, si può assegnare al k-esimo carattere, con

$$\sum_{k=1}^{m} p(k) = 1 \qquad (3.3)$$

L'indice (3.2) è anche noto come indice di Gower (Gordon, 1981).

b) se i caratteri sono ordinali o quantitativi si calcola la distanza tra le unità come la distanza relativa (nel senso chiarito prima) tra le loro posizioni in graduatoria e se ne fa il complemento ad 1; si adotta cioè la formula 1 con la differenza che ora la distanza è una distanza tra le posizioni in graduatoria anziché tra le modalità assunte dalle unità;

c) se fra i caratteri ve ne sono anche di sconnessi, poiché non è possibile calcolare la distanza di Hamming generalizzata tra le unità, si adotta come indice di similarità quello proposto nel *metodo della sintesi di più partizioni* (Zani, 1993). Perciò per ogni carattere si individua una iniziale partizione delle unità e, relativamente al carattere k, si calcola l'indicatore $d_k(i,j)$ che vale 1 se le unità i e j si trovano nello stesso gruppo e vale 0 altrimenti. Tale indice si sostituisce a $V_{ij}(k)$ calcolato per i punti a) e b). In pratica, quindi, la similarità tra due unità risulta essere pari alla frequenza relativa delle partizioni in cui le due unità si trovano incluse in uno stesso gruppo.

Alla matrice delle similarità così ottenuta si applica una procedura di classificazione che è simile alla procedura classica del legame completo secondo la quale una unità può entrare a far parte di un gruppo se la similarità tra essa e tutte le unità di quel gruppo è almeno pari ad un certo livello fissato di similarità. Le fasi di tale procedura sono descritte di seguito.

i) Nella matrice *S* si cercano le eventuali coppie di unità con indice di similarità uguale ad 1: esse formano un gruppo non sfocato detto 'nucleo' (o 'core', Rolland-May, 1985). Il risultato di questa prima fase è quindi una partizione non sfocata perché la funzione di appartenenza non può che assumere i valori caratteristici delle partizioni classiche.
ii) Nella matrice *S* si cerca il massimo valore degli indici di similarità che sia minore di 1. Si individua, così, almeno un indice di similarità *S(i,j)* a cui corrispondono le unità *i* e *j* (le quali formano un nuovo gruppo) e a cui corrisponde il livello di aggregazione α. Si formano tanti gruppi quanti sono gli indici di similarità pari ad α trovati nella matrice *S*. Si determinano i gradi di appartenenza come il minimo delle similarità tra ciascuna unità e tutte le unità comprese nel gruppo.
iii) Si procede in modo analogo al passo precedente: sempre nella matrice *S* si ricerca il massimo valore tra quelli non ancora considerati. Detto $b=S(h,l)$ tale valore, che corrisponde anche al livello di aggregazione, le unità *h* ed *l* si riuniscono in un gruppo. Se però una di esse, ad esempio *h*, risultasse già appartenente ad una gruppo formato in precedenza, si verifica se la similarità tra *l* e tutte le unità del gruppo in cui è compresa *h* non è inferiore pari a *b*: in tal caso *l* può essere assegnata al gruppo a cui già apparteneva *h*, altrimenti le due unità formano un nuovo gruppo. Quando si sono formati tutti i gruppi del livello *b*, si procede alla determinazione della funzione di appartenenza come nel passo precedente.
iv) Si itera il passo iii) per livelli decrescenti degli indici di similarità fino a che non si considerano tutti i possibili livelli.

La procedura di classificazione adottata nel metodo dei ricoprimenti sfocati differisce dalla procedura classica per due ordini di motivi:

- se una unità viene assegnata ad un gruppo essa può ancora essere assegnata ad un altro gruppo a condizione che presenti il necessario livello di similarità con le unità di quel gruppo;
- viene prodotta una *funzione di appartenenza* che fornisce l'insieme dei gradi di appartenenza di ciascuna unità a ciascun gruppo. Il grado di appartenenza di una unità ad un gruppo equivale alla similarità minima tra questa unità e ciascuna delle unità del gruppo.

2.3.2. Un'applicazione

Al fine di valutare più concretamente la capacità descrittiva della metodologia illustrata e per mostrare come si possono analizzare i risultati, si può far riferimento ad una sua applicazione ai dati di mortalità per causa dei 22 paesi europei elencati di seguito:

Tabella 2.3. ***Paesi europei considerati.***

Codice	Nome	Codice	Nome
AUS	Austria	YUG	Iugoslavia
BEL	Belgio	NOR	Norvegia
BUL	Bulgaria	NET	Olanda

CZE	Cecoslovacchia	POL	Polonia
DEN	Danimarca	POR	Portogallo
FIN	Finlandia	FRG	Repubblica Federale Tedesca
FRA	Francia	ROM	Romania
UNK	Gran Bretagna	SPA	Spagna
GRE	Grecia	SWE	Svezia
IRE	Irlanda	SWI	Svizzera
ITA	Italia	HUN	Ungheria

Tabella 2.4. ***Cause di morte considerate.***

Codice	Descrizione
BPI	Influenza, bronchite e polmonite
INF	Malattie infettive ed altre malattie dell'apparato respiratorio
DIG	Malattie dell'apparato digerente
ID	Malattie mal definite
DC	Tumore dell'apparato digerente
BC	Tumore della mammella
UC	Tumore dell'utero
HID	Malattie ischemiche del cuore
RA	Incidenti stradali
SUI	Suicidi
OC	Altri tumori
OCD	Altre malattie cardiovascolari
REM	Restanti cause

In particolare sono stati considerati i tassi di mortalità per le 13 cause di morte descritte nella tabelle 2.4 relativi a donne in età compresa tra 30 e 64 anni nel 1980. I dati sono stati forniti dal World Health Organization (WHO). Abbiamo scelto due livelli di aggregazione: il primo, il livello α (in cui la similarità minima tra le unità è di 0.80), è uno dei livelli più elevati tra tutti quelli che si sono ottenuti e permette di individuare i paesi europei che posseggono una struttura della mortalità molto simile tra loro; il secondo, il livello β (similarità di 0.70), corrisponde al primo livello al quale non si trovano più unità isolate in quanto partecipano tutte alla formazione di almeno un gruppo.

La classificazione sfocata ottenuta ai livelli indicati è riportata di seguito:

Tabella 2.5. ***Classificazione ottenuta al livello di similarità 0.80.***	
GRUPPI	**PAESI EUROPEI**
1	AUS BEL ITA FRG
2	AUS CZE ITA FRG
3	AUS ITA YUG
4	BEL ITA FRG SWE SWI
5	FIN GRE ITA NOR NET SWE SWI
6	FIN GRE ITA SPA
7	FRA ITA SPA
8	UNK IRE
9	ITA NOR FRG SWE SWI
10	YUG POR SPA
Isolati	BUL DEN POL RUM HUN
Fonte: nostre elaborazioni di dati WHO	

Tabella 2.6. ***Classificazione ottenuta al livello di similarità 0.70.***	
GRUPPI	**PAESI EUROPEI**
1	AUS BEL FIN FRA ITA NOR NET FRG SPA SWE SWI
2	AUS BEL CZE FIN UNK ITA NOR FRG SWE
3	AUS BEL FIN UNK ITA NOR FRG SPA SWE
4	AUS BEL CZE FIN ITA NOR FRG SPA SWE SWI
5	BEL FIN FRA GRE ITA NOR NET FRG SPA SWE SWI
6	BEL FIN FRA GRE ITA YUG NOR FRG SPA SWE
7	UNK IRE
8	AUS CZE ITA YUG POR FRG SPA
9	AUS BEL DEN UNK NET FRG SWE
10	AUS CZE HUN
11	BEL FIN UNK ITA NOR NET FRG SWE
12	BUL CZE POL POR
13	CZE DEN UNK FRG SWE
14	CZE ITA YUG POL POR FRG SPA
15	CZE RUM HUN
16	FRA GRE ITA YUG POR FRG SPA
17	YUG NOR POL FRG SPA SWE
18	YUG RUM
Fonte: nostre elaborazioni di dati WHO	

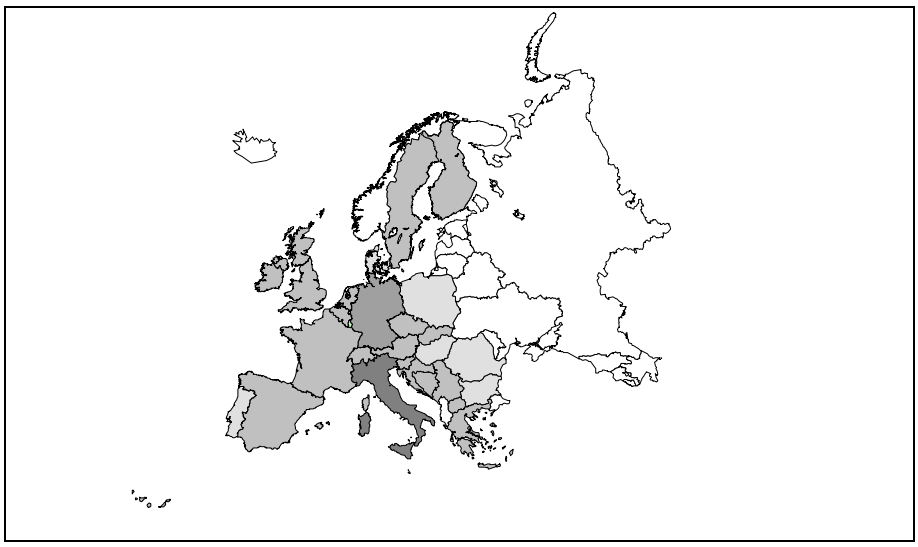
2.3.3. Guida all'analisi dei risultati

Premettiamo che, come comunemente si fa al momento di analizzare i risultati di una classificazione, è sempre possibile calcolare degli indicatori all'interno di ogni gruppo in modo da individuare le caratteristiche proprie del gruppo stesso. La classificazione sfocata offre però altre opportunità che possiamo brevemente descrivere.

In primo luogo possiamo individuare le unità *tipo*, cioè quelle unità che compaiono nei gruppi con la massima frequenza e che quindi hanno una struttura tale che sono simili a buona parte delle unità prese in esame. Nella nostra applicazione, al livello α troviamo che l'Italia è presente i 8 gruppi su 10 (la frequenza relativa delle sue presenze è quindi di 0.8) seguita dalla Germania occidentale che è pre-

sente in 4 gruppi su 10 (con frequenza relativa pari a 0.4). Possiamo perciò ritenere che l'Italia abbia le caratteristiche dell'unità tipo e considerarla come il fulcro attorno al quale si vanno formando i gruppi; è un paese le cui caratteristiche strutturali della mortalità per causa sono molto simili a quelle di altri paesi europei. In particolare, l'Italia presenta caratteristiche strutturali di mortalità per causa molto simili a quelle dell'Europa centrale e settentrionale .

Grafico 2.1. Frequenza con cui i paesi europei compaiono nei gruppi al livello α di similarità.



Queste analogie sono molto recenti. In effetti a partire dal secondo dopo guerra la geografia della mortalità in Europa si è profondamente modificata in seguito alla forte e progressiva diminuzione della mortalità che si è verificata nei paesi del Sud europeo. Una modificazione che ha permesso ai paesi del Sud di raggiungere livelli e caratteristiche di mortalità sempre più simili a quelli dei paesi più sviluppati. Questo comportamento non è stato seguito dai paesi dell'Europa orientale, cioè dai paesi che, anche nella nostra analisi, al livello α rimangono isolati. Nel 1980 l'Europa risulta divisa in due zone contrapposte: da una parte i paesi dell'Europa orientale e dall'altra tutti gli altri paesi. Un risultato questo legato alla recente involuzione sociale, economica e sanitaria che ha interessato l'Est europeo. In questa parte d'Europa, infatti, nel 1980 la mortalità presentava livelli più elevati per le cause (come quelle di natura infettiva e quelle dell'apparato respiratorio e dell'apparato digerente) che più di altre risentono di una scarsa assistenza sanitaria e di condizioni igieniche e socio-economiche meno favorevoli.

Poiché in questo caso i caratteri sono tutti dello stesso tipo e variano tutti nello stesso intervallo (in quanto sono standardizzati), ha senso fare una classificazione in cui i caratteri sono considerati come unità e le unità sono considerate come variabili. Ciò permette di vedere come si raggruppano i caratteri rispetto alle unità e si può rivelare molto utile nell'interpretazione dei risultati poiché si può ritenere che i caratteri che hanno un alto grado di similarità tra loro contribuiscono insieme alla similarità tra le unità. In tal modo la formazione di ciascun gruppo può essere più agevolmente studiata rispetto a gruppi di caratteri anziché rispetto ai caratteri con-

siderati singolarmente. Inoltre i caratteri che tendono a rimanere isolati sono spesso i maggiori responsabili delle differenze tra alcuni gruppi.

Tabella 2.7. ***Classificazione delle cause di morte ottenuta al livello di similarità 0.80.***

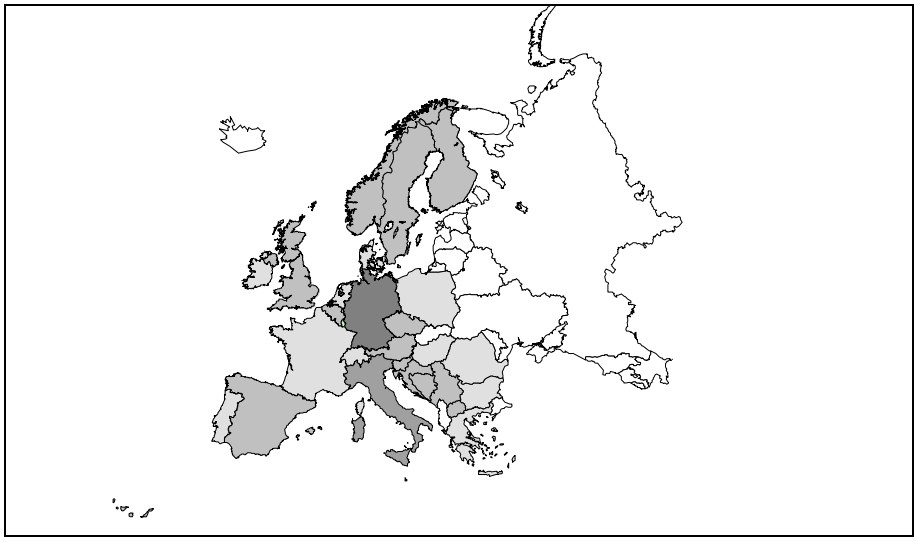
GRUPPI	**CAUSE DI MORTE**
1	DC BC HID OC
2	BPI DC UC HID
3	BPI UC HID OCD
4	INF OCD
5	DIG UC
6	DC UC HID OC
7	BC RA OC
Isolati	ID SUI

Fonte: nostre elaborazioni di dati WHO

Nel nostro caso, le cause di morte di natura infettiva e quelle dell'apparato respiratorio e dell'apparato digerente si aggregano poco con le altre, specialmente con i tumori e con le malattie ischemiche del cuore che sono caratteristiche dei paesi più sviluppati. Le cause di morte che rimangono isolate, cioè le cause di morte mal definite e i suicidi, caratterizzano paesi diversi per struttura di mortalità come, ad esempio, la Francia da una parte e la Romania e la Bulgaria dall'altra.

Quando si considera il livello β, come è ovvio, i gruppi di paesi sono più numerosi e, in questo caso, è la Germania occidentale a diventare unità tipo, essendo presente in 13 gruppi su 18 (frequenza relativa di 0.7) seguita dall'Italia con 10 presenze su 18 (frequenza relativa 0.6).

Grafico 2.2. Frequenza con cui i paesi europei compaiono nei gruppi al livello β di similarità.



Lo scambio di posizioni tra i due paesi, passando dal livello α al livello β, potrebbe essere determinato dal fatto che al livello più basso si aggregano anche i paesi dell'Est europeo sicuramente più simili alla Germania che all'Italia. Infatti, al livello β quest'ultima si aggrega ancora una volta solo con i paesi dell'Europa occi-

dentale. Inoltre a questo livello anche la Danimarca, che al livello precedente rimaneva isolata, entra a far parte di alcuni paesi dell'Europa occidentale. In particolare appare simile all'Austria, Belgio, Germania occidentale e Olanda. È interessante vedere quali siano le cause di morte che permettono alla Danimarca di aggregarsi con i paesi indicati. Per fare ciò costruiamo una tabella che ha tante righe quanti sono i caratteri e tante colonne quante sono le coppie di unità da esaminare. Ogni riga contiene gli indici parziali di similarità, cioè i valori $V_{ij}(k)$. In questo modo si può individuare il contributo di ogni carattere alla formazione di una coppia di unità contenuta in un gruppo.

Tabella 2.8. ***Similarità parziali tra alcuni paesi.***

Causa di morte	DEN-AUS	DEN-BEL	DEN-NET	DEN-FRG
BPI	0.58	0.66	0.41	0.62
INF	**0.94**	**0.93**	**0.88**	**0.99**
DIG	0.53	0.85	0.86	0.54
ID	0.53	0.51	0.95	0.73
DC	0.93	0.60	0.63	0.89
BC	0.66	1.00	0.92	0.66
UC	0.97	0.64	0.58	0.72
HID	0.86	0.81	0.91	0.98
RA	0.77	0.55	0.88	0.99
SUI	**0.45**	**0.62**	**0.26**	**0.45**
OC	**0.42**	**0.35**	**0.30**	**0.49**
OCD	**0.99**	**0.96**	**0.92**	**0.98**
REM	0.94	0.61	0.84	0.96

***Fonte**:* nostre elaborazioni di dati WHO

Le righe evidenziate di questa tabella mostrano che sono le malattie infettive e le malattie cardiovascolari le principali cause di morte responsabili di queste aggregazioni, mentre i suicidi e gli altri tumori contribuiscono a differenziare i paesi considerati.

2.4.1. Il metodo del legame medio sfocato

Anche questo metodo è di tipo gerarchico e produce partizioni sfocate delle unità considerate. La procedura di classificazione si suddivide in due momenti principali: il calcolo di una matrice di similarità e la procedura aggregativa vera e propria.

L'indicatore di similarità adottato è quello descritto nel paragrafo 2.3.1 punto a), cioè si tratta di un metodo adatto allo studio di casi in cui si abbiano caratteri di tipo quantitativo.

S(i,j) è un indice di similarità a cui può essere applicata la procedura di classificazione seguente:

Passo 1: Si cercano le coppie di unità che presentano il valore massimo dell'indice di similarità (può verificarsi il caso che tali coppie siano più di una). Tali coppie formano i gruppi iniziali (nuclei). Si calcolano i centroidi relativi ad ogni gruppo: ogni centroide rappresenta un nuovo individuo avente caratteristiche intermedie agli individui presenti nel gruppo e quindi si può considerare rappresentativo del gruppo stesso. Si calcolano quindi i gradi di appartenenza di ciascuna unità

ad ogni gruppo in proporzione inversa alla distanza di tale unità dal centroide del gruppo (più l'unità è lontana dal centroide, minore sarà il suo grado di appartenenza al gruppo). La somma dei gradi di appartenenza di ciascuna unità ad ogni gruppo deve essere uguale ad 1 (condizione necessaria perché si abbia una partizione sfocata anziché un ricoprimento), per cui se l'unità appartiene ad un solo gruppo il suo grado di appartenenza è uguale ad 1.

Passo 2: Si considera ancora la matrice *S(i,j)* nella quale si cercano le coppie di unità che presentano il valore della similarità più alto (escludendo i valori già considerati). Sia α tale valore della similarità. Le coppie individuate possono essere composte di unità che erano già state considerate al passo precedente. Per semplicità, supponiamo di aver individuato la coppia *(i', j')* e supponiamo che l'unità *i'* era già stata inserita in un gruppo; in tal caso *j'* può essere inserita nel gruppo a cui appartiene *i'* solo se la similarità media tra *j'* e tutte le unità che compongono il gruppo è maggiore di α. Altrimenti le unità *i'* e *j'* formano un nuovo gruppo. Una volta formati tutti i gruppi, si calcolano i centroidi e gradi di appartenenza come specificato nel passo 1.

Passo 3: Si itera il passo 2 fino a quando tutte le unità formano un unico gruppo.

2.4.2. Un'applicazione

Il metodo del legame medio sfocato è stato applicato in molti casi. A titolo d'esempio, possiamo mostrare una applicazione fatta come analisi di approfondimento in occasione di un'indagine condotta dall'IRP nel 1997 sulle opinioni degli italiani riguardo le recenti tendenze demografiche.

I dati dell'indagine sono stati preventivamente trasformati in quantitativi utilizzando alcune variabili strutturali scelte tra quelle disponibili: gli individui che presentavano le stesse modalità di queste variabili sono stati uniti nello stesso gruppo ed hanno costituito una **tipologia semplice (o strato)**. Ovviamente, la selezione di queste variabili ha richiesto uno studio preliminare dei dati perché occorreva selezionare strati di un qualche interesse, ma che nello stesso tempo non fossero troppo numerosi e contenessero un numero di individui intervistati non troppo diverso tra loro. Le variabili scelte per l'individuazione delle tipologie semplici sono quelle che erano state considerate in fase di piano del campionamento, oltre all'informazione sul numero di figli che l'intervistato aveva già avuto. Cioè sono le seguenti tre variabili:

1) Sesso, costituita da 2 modalità: maschio e femmina.
2) Età, di tre modalità: ventenni (20-29 anni), trentenni (30-39 anni) e quarantenni (40-59 anni).
3) Numero di figli, costituita da 3 modalità: nessuno, uno, due o più.

Per l'esigua numerosità degli intervistati ventenni con due o più figli, solo per questa classe d'età si sono considerate due modalità della terza variabile strutturale: nessuno e uno o più.

Si calcolano quindi le frequenze con cui questi gruppi di individui hanno risposto alle domande. In tal modo è possibile operare come se i dati fossero quantitativi e il numero di unità da classificare è notevolmente ridotto rispetto al numero di intervistati.

La combinazione delle modalità di queste quattro variabili strutturali dà luogo in definitiva a 16 tipologie di individui (tabella 2.9). Agli strati, considerati come unità, e alle frequenze con cui questi gruppi di individui hanno risposto alle domande, considerate come variabili, è stata applicata la procedura di classificazione del metodo del legame medio sfocato. In questa applicazione però, per motivi di brevità, vedremo la classificazione riferita ad uno solo degli argomenti trattati nell'indagine, cioè la parte relativa alle *intenzioni riproduttive*. La classificazione ottenuta è riportata nella tabella 2.10.

Se si calcolano le medie delle variabili utilizzate ponderate con i gradi di libertà in ogni gruppo e si confrontano con le medie generali, si ottengono facilmente le interpretazioni dei gruppi che si sono formati.

Gruppo 1: Sono persone piuttosto giovani che non hanno ancora avuto figli ma che sono intenzionate ad averne. Dalle risposte che danno, però non sembra che abbiano le idee ben chiare sul valore dei figli e sul costo che comporta averne, probabilmente perché non ne hanno l'esperienza.

Gruppo 2: Sono accomunati dal fatto di avere già il numero di figli che desideravano per cui non intendono averne ancora. Apprezzano senza dubbio il valore che un figlio può avere, ma i costi e la mancanza di servizi sociali di supporto ai genitori li rendono sicuri di non volere altri figli nel futuro.

Gruppo 3: Sono gli indecisi: hanno già dei figli ma si sentono abbastanza giovani per averne ancora anche se sono preoccupati per i costi che ciò comporta e per la scarsezza dei servizi sociali per l'infanzia.

Gruppo 4: Non intendono avere figli perché pensano di averne abbastanza, anche se sono convinti che i figli diano un valore aggiuntivo alla loro esistenza. Quando si chiede loro se sono soddisfatti dei servizi per l'infanzia, pensano che manchino asili nido, ma non sono preoccupati per i servizi di sorveglianza prima e dopo la scuola o durante le vacanze (come se la cosa non li riguardasse in quanto sanno a chi affidare i bambini in quel periodo).

Gruppo 5: Sono donne che ritengono di non essere più nella stagione della maternità, sia per l'età, sia perché hanno già abbastanza figli. Rimangono comunque preoccupate per la mancanza di servizi (di tutti i tipi) per l'infanzia perché temono che essi rappresentino un freno per i genitori che intendono avere altri figli, mentre per loro la maternità è stata l'esperienza più importante e gratificante della vita.

Isolati: Lo strato numero 2 sembra essere quello dei 'disinteressati' al problema: non danno importanza all'assenza dei servizi per l'infanzia e non ritengono che gli impegni di lavoro siano di intralcio famiglie con figli, tuttavia pensano che in futuro avranno altri figli.

Lo strato 6 è composto da uomini che sembrano non essersi posti il problema di avere o no figli nella loro vita, ma pensano che ormai sono troppo avanti con l'età per averne. In via teorica ritengono che sia importante per le famiglie poter contare sui servizi per l'infanzia e che i figli sono una cosa importante ma non fondamentale.

Lo strato 14 è composto da donne certe di non volere figli nel futuro per motivi di salute e perché non lo ritengono necessario in quanto ci si può realizzare anche in altre attività della vita oltre la maternità.

Tabella 2.9. ***Strati considerati.***

Strato	Sesso	Età	N. figli
1	Maschio	Ventenni (20/29)	Nessuno
2	Maschio	Ventenni (20/29)	Uno o più
3	Maschio	Trentenni (30/39)	Nessuno
4	Maschio	Trentenni (30/39)	Uno
5	Maschio	Trentenni (30/39)	Due o più
6	Maschio	Quarantenni (40/49)	Nessuno
7	Maschio	Quarantenni (40/49)	Uno
8	Maschio	Quarantenni (40/49)	Due o più
9	Femmina	Ventenni (20/29)	Nessuno
10	Femmina	Ventenni (20/29)	Uno o più
11	Femmina	Trentenni (30/39)	Nessuno
12	Femmina	Trentenni (30/39)	Uno
13	Femmina	Trentenni (30/39)	Due o più
14	Femmina	Quarantenni (40/49)	Nessuno
15	Femmina	Quarantenni (40/49)	Uno
16	Femmina	Quarantenni (40/49)	Due o più

Tabella 2.10. ***Classificazione sfocata ottenuta al livello di similarità media 0.78.***

Gruppo	Num. strato	Descrizione			Gda
1	1	Maschio	Ventenni (20/29)	Nessuno	1
	3	Maschio	Trentenni (30/39)	Nessuno	1
	9	Femmina	Ventenni (20/29)	Nessuno	1
	11	Femmina	Trentenni (30/39)	Nessuno	1
2	8	Maschio	Quarantenni (40/49)	Due o più	0.4
	13	Femmina	Trentenni (30/39)	Due o più	1
	16	Femmina	Quarantenni (40/49)	Due o più	0.59
3	4	Maschio	Trentenni (30/39)	Uno	1
	5	Maschio	Trentenni (30/39)	Due o più	1
	8	Maschio	Quarantenni (40/49)	Due o più	0.23
	10	Femmina	Ventenni (20/29)	Uno o più	1
	12	Femmina	Trentenni (30/39)	Uno	1
4	7	Maschio	Quarantenni (40/49)	Uno	1
	8	Maschio	Quarantenni (40/49)	Due o più	0.37
5	15	Femmina	Quarantenni (40/49)	Uno	1
	16	Femmina	Quarantenni (40/49)	Due o più	0.41
Isolati	2	Maschio	Ventenni (20/29)	Uno o più	
	6	Maschio	Quarantenni (40/49)	Nessuno	
	14	Femmina	Quarantenni (40/49)	Nessuno	

Fonte*:* nostre elaborazioni di dati IRP

Notiamo, infine, che lo strato 8 e lo strato 16 si ritrovano in più di un gruppo in quanto, evidentemente, le persone che li compongono hanno comportamenti rispetto alle intenzioni riproduttive che non possono essere inglobati in un solo schema.

2.5. Osservazioni

Tutti i metodi qui proposti si basano sul calcolo di una matrice di indici di similarità tra le coppie di unità. Si può presentare quindi il problema di applicare tali

metodi a dati di qualsiasi ordine di misura. In genere, infatti, è agevole calcolare distanze e similarità soltanto se si considerano caratteri quantitativi o, al più, ordinali; in realtà sono state proposte anche diverse soluzioni di calcolo delle similarità per caratteri qualitativi, (ad esempio, A. Di Ciaccio, 1990). L'indice di similarità proposto da S. Zani risolve con estrema semplicità il problema del calcolo degli indici di similarità per caratteri di qualunque ordine di misura. Tale indice, può non risultare il migliore quando si dispone soltanto di caratteri quantitativi situazione che nelle applicazioni pratiche è forse la più frequente.

Infatti, se per i caratteri qualitativi la scelta della partizione iniziale non è un problema rilevante in quanto la partizione migliore è proprio quella individuata dalle stesse modalità assunte dalle unità, nel valutare la similarità per caratteri quantitativi c'è, invece, un più ampio margine di soggettività al momento di individuare la partizione iniziale delle unità. Questo modo di procedere ha comunque il vantaggio di poter calcolare un indice di similarità che sintetizzi l'informazione fornita da caratteri di tipo diverso. Ma se si hanno caratteri tutti di uno stesso tipo, ad esempio solo caratteri quantitativi, oppure solo caratteri rettilinei ordinati, con questo indice di similarità i dati vengono trattati nello stesso modo in cui si tratterebbero se si avessero solo caratteri sconnessi. Ciò può risultare poco conveniente nei casi in cui è possibile usare un indice di similarità che si adatti di più alla natura dei caratteri considerati.

Da qui nasce l'idea di considerare un indicatore di similarità che si adatti al minimo livello di misura dei dati poiché è ovvio che è meglio usare un indice che sfrutta al massimo l'informazione disponibile. Inoltre, buona parte delle situazioni reali fa riferimento a caratteri quantitativi o che si possono rendere tali con facilità (ad esempio è il caso delle frequenze o delle percentuali che possono far riferimento a fenomeni per loro natura qualitativa ma che di fatto possono essere analizzati con metodi adatti a dati quantitativi) per cui è buona norma prevedere sempre un indice si similarità adatto a caratteri quantitativi.

Per quanto riguarda la procedura di classificazione, occorre chiarire che il metodo usato da Zani differisce dal metodo classico del legame singolo per il fatto che per quest'ultimo, se una unità entra a far parte di un gruppo, essa non può più essere assegnata ad altri gruppi e quindi in seguito non verrà più presa in considerazione. Il metodo di Zani, invece, è in metodo di classificazione sfocata per cui, se una unità entra a far parte di un gruppo con grado di appartenenza minore di 1, essa potrà ancora essere presa in considerazione perché può appartenere, per la parte rimanente, ad altri gruppi. In tal modo, però, si può intensificare il noto ***effetto catena***, tipico del metodo del legame singolo, che produce la formazione di gruppi in cui le unità possono non essere tutte simili allo stesso livello, ma fanno ugualmente parte di quel gruppo perché sono simili ad almeno una delle unità del gruppo. Questo effetto si amplifica se si adotta un metodo di classificazione sfocata in quanto le unità vengono considerate più di una volta e quindi aumenta il rischio di associarle a gruppi in cui sono già presenti unità dissimili da esse.

Questo implica che è stata tacitamente adottata la proprietà di transitività max-min secondo cui, una unità è simile ad un livello α alle unità inserite in un gruppo, se il massimo dei gradi di similarità tra essa e ognuna delle altre unità è almeno uguale ad α. Ma se vogliamo che valga una proprietà di transitività più forte che garantisca che la similarità delle unità appartenenti allo stesso gruppo sia almeno

uguale ad un certo livello, dobbiamo adottare un metodo simile al metodo classico del legame completo: una unità è inserita in un gruppo al livello α con un certo grado di appartenenza se essa presenta grado di similarità con tutte le unità già presenti nel gruppo, almeno pari ad α (come nel *metodo dei ricoprimenti sfocati*). Cioè si adotta una transitività di tipo min-max, anche detta *proprietà di affinità sfocata* (L.A. Zadeh, 1975). Ciò evita che si formino gruppi con unità poco simili tra loro anche se il numero di gruppi che si formano può essere molto superiore al numero di gruppi che si ottenevano, allo stesso livello, con la procedura di classificazione proposta da Zani, ma in compenso è garantita una maggiore omogeneità tra le unità che fanno parte di uno stesso gruppo.

Naturalmente, una classificazione che rispetta la proprietà di affinità sfocata può risultare piuttosto complessa. Questa complessità, da un lato può essere limitata, come si è detto, con delle semplici scelte per selezionare i risultati ottenuti, e dall'altro comporta una quantità notevole di informazione che con gli altri metodi va quasi sempre persa.

Un'alternativa al *metodo della sintesi di più partizioni* e al *metodo dei ricoprimenti sfocati* può essere rappresentata dal *metodo del legame medio sfocato*, ancora una volta di ispirazione classica (nel senso che è simile al metodo del legame medio classico), e che fornisce gruppi formati da unità abbastanza omogenee tra loro in quanto perché una unità entri a far parte di un gruppo occorre che la similarità media tra essa e tutte le altre unità già presenti nel gruppo sia almeno pari ad una certa soglia α, ma il numero di gruppi che si forma è più o meno intermedio al numero di gruppi che si formano con gli altri due metodi e ai livelli più bassi della gerarchia la sfocatura diventa talmente alta che le funzioni di appartenenza di ciascuna unità tendono ad equidistribuirsi tra i gruppi.

Osserviamo ora che il vincolo secondo cui la somma dei gradi di appartenenza di ogni unità ai gruppi deve valere 1 (che da ora in poi chiameremo ***vincolo della somma per riga***) può impedire la formazione di alcuni gruppi qualora sia imposto prima che sia stata completata la procedura di aggregazione. E opportuno introdurre tale vincolo a posteriori rispetto alla formazione dei gruppi, attraverso, ad esempio, una normalizzazione dei gradi di appartenenza, cioè dividendo i gradi di appartenenza di ciascuna unità per la somma dei gradi di appartenenza di quella unità a tutti i gruppi a cui essa appartiene. In realtà, tale normalizzazione può anche non avere luogo in quanto ha l'effetto di diminuire alcuni gradi di appartenenza e quindi di fornire una lettura dei risultati non del tutto realistica. Se non si effettua la normalizzazione anziché ottenere delle partizioni sfocate si ottengono dei ricoprimenti sfocati. E' ovvio, quindi, che i risultati dei metodi presentati in questo capitolo possono essere alternativamente partizioni o ricoprimenti a seconda che si imponga o no il vincolo della somma per riga senza che questo cambi nella sostanza le aggregazioni che si ottengono a patto che tale vincolo venga imposto alla fine della procedura di aggregazione.

Capitolo 3

METODI NON GERARCHICI DI CLASSIFICAZIONE SFOCATA

di Giampaolo Iacovacci

3.1. Introduzione

I metodi di classificazione sfocata esaminati in questo capitolo sono caratterizzati dall'essere tutti di tipo non gerarchico, ossia ciascuno di essi fornisce una classificazione sfocata delle unità in un ben determinato numero di gruppi che viene stabilito a priori fin dall'inizio della procedura di classificazione.

La classificazione viene poi ottenuta attraverso un processo iterativo tendente alla ottimizzazione di una funzione obiettivo che, in genere, rappresenta una misura della dispersione dei punti dai centri dei cluster.

La differenza principale tra i metodi di questo tipo consiste di fatto nella diversa funzione obiettivo adottata e, dunque, nel differente processo iterativo utilizzato per calcolare i gradi di appartenenza delle unità ai vari gruppi.

Nel seguito saranno esposti i metodi di classificazione più noti ma, essendo senza dubbio il metodo delle k-medie sfocato quello più conosciuto ed utilizzato, ad esso è dedicata una maggiore attenzione rispetto agli altri.

Nei paragrafi 3.3. e seguenti sono illustrati alcuni recenti perfezionamenti del metodo delle k-medie sfocato e l'applicazione di uno di questi per la classificazione dei comuni urbani e rurali.

Altri metodi di classificazione sono brevemente descritti nel paragrafo 3.4. mentre il paragrafo 3.5. contiene alcune osservazioni finali.

3.2. Il metodo delle k-medie sfocato

Tra i metodi di classificazione sfocata di tipo non gerarchico, il più conosciuto ed utilizzato è senz'altro il metodo delle k-medie sfocato (Bezdek, 1981).

Esso rappresenta una generalizzazione del metodo classico delle k-medie ed è particolarmente indicato per trattare grosse matrici di dati poiché la convergenza verso la classificazione finale viene generalmente raggiunta in breve tempo.

Per utilizzare il metodo delle k-medie sfocato si procede nel seguente modo:

dopo aver scelto il numero c di cluster in cui si vogliono suddividere le n unità sulle quali sono state rilevate le modalità x di p caratteri, si fornisce una partizione iniziale delle unità nei c gruppi (che può essere casuale o basata su conoscenze a priori del ricercatore). Partendo da questa si ottiene, attraverso successive iterazioni tendenti alla minimizzazione di una funzione obiettivo, una classificazione sfocata nella quale per ogni unità viene determinato il grado di appartenenza ai c gruppi.

Il grado di appartenenza di un'unità ad un gruppo viene espresso per mezzo dei valori μ_{ik} assunti dalla funzione di appartenenza i quali sottostanno ai seguenti vincoli:

$$\text{i)} \quad 0 \le \mu_{ik} \le 1 \qquad i=1,...,n \quad k=1,...,c$$

ii) $\sum_{k=1}^{c} m_{ik} = 1$ i=1,...,n

dove il vincolo i) definisce l'insieme di definizione della funzione di appartenenza ed il vincolo ii) indica che la somma dei gradi di appartenenza di ogni unità ai vari gruppi deve valere 1.

L'insieme dei valori della funzione di appartenenza può essere rappresentato in una matrice $U = [\mu_{ik}]$ di dimensione (n x c).

La funzione obiettivo da minimizzare, detta *funzione di ottimizzazione* J_m, viene utilizzata per calcolare i valori ottimi del grado di appartenenza ed è funzione del quadrato della distanza d_{ik} tra l'unità i-esima ed il centroide del k-esimo cluster e dipende da un parametro m che può assumere qualunque valore reale ≥ 1:

$$J_m(U,v) = \sum_{k=1}^{c}\sum_{i=1}^{n}(m_{ik})^m (d_{ik})^2 \qquad (3.1)$$

dove $(d_{ik})^2 = |x_i - v_k|^2$ e $|\cdot|$ è un'opportuna norma su R^p;
$v_k \in R^p$ è la componente k-esima del vettore dei centroidi $v = (v_1,..,v_c) \in R^{cp}$;
$x_i \in R^p$ è la componente i-esima del vettore delle unità $x = (x_1,..,x_n) \in R^{np}$;
l'esponente $m \in [1,\infty)$;
U è la matrice di dimensione (n x c) dei gradi di appartenenza.

La funzione obiettivo J_m ha una chiara interpretazione: per ogni dato cluster, il suo centroide è la miglior rappresentazione delle unità che lo compongono poiché esso minimizza la somma dei quadrati degli errori $x_i - v_k$. Così, J_m misura l'errore quadratico totale in cui si incorre nel rappresentare le n unità con i c centroidi dei cluster. Il valore di J_m dipende allora da come le unità sono raggruppate nei cluster e rappresenta dunque una misura della dispersione delle unità intorno ai centri dei cluster; la partizione ottima è considerata quella che minimizza J_m. Tale partizione è anche chiamata partizione di minima varianza.

Il parametro m che compare nella (3.1) riveste una particolare importanza poiché, a seconda del valore che si sceglie, valore che deve essere fornito all'inizio della procedura, la classificazione che si otterrà sarà più o meno sfocata (maggiori approfondimenti sul parametro m sono forniti nei paragrafi 3.2.3. e 3.2.4.).

L'algoritmo che descrive il metodo delle k-medie sfocato è il seguente:

Passo 1: fissato il valore di $m \in [1,\infty)$ e di $c \in [2,n)$ e scelta la metrica da utilizzare, si sceglie una partizione iniziale delle unità in c gruppi che può essere rappresentata con la matrice $U^{(0)} = [\mu_{ik}]$ dove con l'esponente si indica il numero di iterazioni.

Passo 2: si calcolano i c centroidi dei cluster $v_k^{(0)}$ usando la formula:

$$V_k^{(0)} = \frac{\sum_{i=1}^{n}(m_{ik})^m x_i}{\sum_{i=1}^{n}(m_{ik})^m} \qquad (3.2)$$

Passo 3: si calcola la nuova matrice $U^{(1)}$, che rappresenta il risultato della prima iterazione, secondo le seguenti regole:

a) se per qualche gruppo r si ha che $d_{ir}= 0$, si pone $\mu_{ir}= 1$ e $\mu_{ik}= 0$ per tutti i $k \neq r$;

b) se la precedente condizione non è soddisfatta allora si utilizza la seguente formula:

$$m_{ik} = \frac{1}{\sum_{j=1}^{c} \left[\frac{d_{ik}}{d_{jk}} \right]^{2/(m-1)}} \qquad (3.3)$$

Passo 4: si calcola la differenza tra i risultati ottenuti all'ultima e alla penultima iterazione usando un'opportuna norma: se

$$| U^{(1)} - U^{(0)} | \leq \delta , \qquad (3.4)$$

dove δ è un parametro stabilito a priori, allora ci si ferma e si considera come classificazione finale quella ottenuta all'ultima iterazione, altrimenti si torna al passo 2 e si esegue una nuova iterazione continuando il procedimento fin quando la (3.4) non è soddisfatta.

Nei prossimi paragrafi verranno esaminate le principali proprietà di cui gode il metodo delle k-medie sfocato.

3.2.1. La convergenza

Uno dei principali motivi per cui il metodo delle k-medie sfocato è molto utilizzato, risiede nella rapidità con la quale tale metodo arriva alla classificazione finale.

Notiamo inoltre che studi recenti (Bezdek e Hathaway, 1988) hanno dimostrato che le soluzioni finali ottenute corrispondono sempre e solamente ad un punto di minimo locale o globale della funzione obiettivo (3.1) o, al peggio, ad un suo punto sella. D'altra parte, ciò può non essere uno svantaggio, poiché, seppure un punto di minimo globale è senza dubbio da preferire ad uno di minimo locale o ad un punto sella, è stato anche fatto notare che, in molti casi, le classificazioni corrispondenti ai tre diversi punti sono praticamente identiche tra loro.

Un'altra delle caratteristiche positive della convergenza del metodo delle k-medie sfocato, consiste nel fatto che ad ogni successiva iterazione il valore della funzione obiettivo decresce rispetto a quello dell'iterazione precedente cosa che invece non sempre si verifica negli altri metodi di questo tipo suscitando così non poche perplessità sulla loro convergenza.

Infine, resta da segnalare che osservazioni empiriche (vedi esempio del paragrafo 3.3.3.) hanno messo in evidenza che il metodo risulta essere "relativamente indipendente dalla scelta della partizione iniziale" (Bezdek e Hathaway, 1988) convergendo comunque sempre allo stesso punto della funzione obiettivo.

3.2.2. La scelta della partizione iniziale $U^{(0)}$ e del numero c di cluster

Come già detto nel paragrafo precedente, nella maggior parte dei casi il metodo delle k-medie sfocato fornisce lo stesso risultato indipendentemente da quale sia la partizione iniziale che può essere così casuale o scelta dall'utilizzatore in base a

precedenti conoscenze del fenomeno. È ovvio però che, quanto più la partizione di partenza si avvicina a quella finale, tanto più si accelera il processo di convergenza. Dunque, per ridurre i tempi di elaborazione, sembra essere buona regola scegliere sempre come $U^{(0)}$ quella ottenuta mediante un altro metodo di classificazione (classico o sfocato).

Per quanto riguarda la scelta del numero c di cluster, questa è lasciata di fatto alla sensibilità del ricercatore che dovrà basarsi in generale sulle supposizioni o conoscenze che ha del fenomeno. Infatti, poiché da uno studio dei diversi test proposti per la scelta del numero di cluster (Milligan,1985) risulta che nessuno di questi può essere considerato esente da difetti, non esiste un criterio oggettivo per scegliere il parametro c. L'unico procedimento che può essere adoperato per aiutarsi nella scelta, è quello di provare alcuni valori di c e di confrontare le classificazioni ottenute facendo però attenzione a non scegliere un numero di cluster troppo grande poiché, essendo ogni unità ripartita in parte nei vari cluster, si potrebbe ottenere una classificazione troppo sfocata e dunque di difficile interpretazione.

3.2.3. La scelta del parametro m

All'inizio della procedura di classificazione si deve scegliere, tra gli altri parametri, il valore di m. Questa scelta riveste una particolare importanza poiché a seconda del valore di m dato, la classificazione che si otterrà sarà più o meno sfocata. In generale, si possono verificare due casi:

a) **m=1**; se viene scelto questo valore per il parametro m, si dimostra che la classificazione delle unità è di tipo classico, ossia il grado di appartenenza assume solamente i valori 0 o 1 eliminando così qualunque tipo di sfocatura. In tal caso il metodo delle k-medie sfocato coincide con il metodo delle k-medie classico, che risulta essere così un caso particolare. Ogni unità viene in questo caso attribuita totalmente al cluster da cui ha distanza minore.
b) **m > 1** ; in questo caso, quanto più il valore di m sarà maggiore di 1, tanto più il grado di appartenenza tenderà ad assumere, per ogni unità, valori sempre più lontani dagli estremi 0 e 1 fino ad ottenere, al limite, il valore 1/c corrispondente al caso di massima sfocatura della classificazione, dove ogni unità è equamente distribuita tra tutti i c cluster presenti.

Come è evidente, per riuscire ad ottenere un buon risultato la scelta di m risulta così determinante. Spesso, però, in assenza di ulteriori informazioni sul fenomeno da analizzare, tale scelta è piuttosto complicata e, per risolvere tale problema, sono state proposte alcune soluzioni, una delle quali verrà esposta nel seguente paragrafo.

3.2.4. Una soluzione per la scelta di m

Le diverse applicazioni effettuate con il metodo delle k-medie sfocato hanno messo in evidenza che non esiste un valore ottimale per il parametro m. Esso infatti varia a seconda delle singole applicazioni e ciò rappresenta un notevole difetto del metodo poiché la scelta di m può essere fatta solamente osservando le classificazioni finali ottenute in corrispondenza dei diversi valori e, in ogni caso, se non si

hanno conoscenze del fenomeno indagato risulta difficile scegliere il livello di sfocatura ideale delle classificazioni.

Per ovviare a questo inconveniente, si è pensato di introdurre un indice che misuri il grado di sfocatura delle diverse classificazioni tenendo conto che una classificazione si dice totalmente sfocata se, detto c il numero di cluster, per ogni unità tutti i valori della funzione di appartenenza assumono valore 1/c, oppure si dice classica se ogni unità appartiene ad un unico cluster.

Si noti che, secondo queste definizioni, il concetto di sfocatura è equivalente a quello di eterogeneità, cosicché una classificazione può essere detta più o meno sfocata a seconda se essa sia più o meno eterogenea. Da quanto detto, segue che un qualunque indice di eterogeneità può anche essere considerato un indice di sfocatura per cui per misurare il grado di sfocatura delle classificazioni si propone di utilizzare l'indice relativo di eterogeneità di Gini:

$$I = \frac{c}{c-1}\left[1 - \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{c} m_{ik}^{2}\right] \qquad (3.5)$$

Tale indice varia nell'intervallo [0,1] ed assume valore 1 nel caso di massima sfocatura e valore 0 nel caso in cui la classificazione è di tipo classico. Con l'aiuto di questo indice è così possibile avere una misura sintetica dell'effetto dei diversi valori di m sulle classificazioni corrispondenti e la scelta di m risulta così agevolata.

Dalle diverse applicazioni effettuate il valore migliore di m sembra essere quello ottenuto in corrispondenza del valore dell'indice I compreso tra 0.4 e 0.5.

3.3. Perfezionamenti del metodo delle k-medie sfocato

Nei paragrafi precedenti è stato sottolineato come il metodo delle k-medie sfocato, grazie alle proprietà di cui gode, risulti uno dei migliori metodi di classificazione sfocata.

Studi recenti hanno però evidenziato che, alcune volte, questo metodo può fornire una classificazione eccessivamente sfocata; in particolare, si considerino i due seguenti casi:

a) se il numero c di cluster è abbastanza grande, il fatto che ogni unità debba appartenere (almeno in parte) ad ognuno dei cluster, fa sì che la classificazione stessa potrebbe risultare troppo "frastagliata" e, dunque, di difficile interpretazione;
b) se le unità sono raggruppate in cluster ben netti e separati tra loro, qualunque sfocatura nella classificazione risulterebbe fornire un risultato distorto.

Per eliminare alcuni di questi inconvenienti, recentemente sono stati proposti dei suoi perfezionamenti che danno luogo ad una classificazione che chiameremo **classificazione semisfocata** per distinguerla da quella sfocata prodotta dal metodo delle k-medie sfocato.

Questi nuovi metodi saranno presentati nei prossimi due paragrafi.

3.3.1. I metodi di Kamel, Selim e Ismail

Nel 1984 Selim e Ismail hanno proposto tre nuovi metodi di classificazione semisfocata corrispondenti ad altrettante variazioni del metodo delle k-medie sfocato che, a parte queste, rimane inalterato. Tali variazioni sono le seguenti:

1) oltre al numero dei c cluster iniziali, si sceglie anche un numero di cluster $\pi < c$ e si impone che ogni unità possa appartenere al più a π cluster; ciò si ottiene ordinando, per ogni unità, in ordine decrescente i primi π valori della funzione di appartenenza e ponendo uguale a zero i restanti $c-\pi$ valori.
   Per rispettare il vincolo secondo il quale per ogni unità la somma dei gradi di appartenenza deve essere uguale a 1, si opera poi la seguente normalizzazione:

$$w_{ik} = \frac{m_{ik}}{\sum_{k=1}^{c} m_{ik}} \qquad (3.6)$$

   dove con w_{ik} si è indicato il valore della funzione di appartenenza normalizzata dell'unità i-esima al k-esimo cluster;
2) si impone che se un'unità è molto distante da un certo cluster, il suo grado di appartenenza a quel cluster è nullo, dunque, detta d_{ij} la distanza tra l'unità i-esima e il centro del j-esimo cluster, se $d_{ij} > \gamma$ (dove γ è un valore prefissato), si pone $\mu_{ij} = 0$. Anche in questo caso si procede poi alla normalizzazione dei valori μ_{ij} mediante la (3.6).
3) per eliminare la presenza di valori molto bassi della funzione di appartenenza, e rendere così più chiara la classificazione finale, se $\mu_{ij} < ß$ (dove ß è un valore scelto a priori), si pone $\mu_{ij} = 0$.
   Anche qui, come nei due metodi precedenti, si procede poi alla normalizzazione.

I tre metodi proposti, come gli stessi autori ammettono, non riescono però a risolvere tutti i problemi connessi con il metodo delle k-medie sfocato:

il metodo 1 infatti, risulta utile applicarlo solamente quando si possiedono precise informazioni a priori per poter scegliere un idoneo numero π di cluster, altrimenti limitare le unità ad appartenere ad un numero massimo di cluster può portare a serie distorsioni nel calcolo globale dei valori del grado di appartenenza delle unità stesse.

Il metodo 2 è utile se si ha il sospetto che esistano diversi dati anomali, ma anche in questo caso bisogna fare attenzione alla scelta del valore γ perché se tale parametro viene scelto troppo piccolo, alcune unità potrebbero non essere assegnate ad alcun cluster.

Il metodo 3, infine, è quello che ha riscosso più consensi, in quanto sembra essere di più generale utilità e, inoltre, facendo una scelta appropriata del parametro ß, per ogni unità viene così determinato in modo naturale il numero di cluster al quale appartiene, superando così il problema della scelta a priori del numero di cluster π incontrato nel metodo 1. Anche la scelta del parametro ß però, non sempre risulta semplice, in quanto se viene scelto un valore troppo alto, alcune unità potrebbero non appartenere ad alcun cluster o, addirittura, qualche cluster potrebbe risultare vuoto.

Per superare il problema della scelta del parametro ß, nel 1991 Kamel e Selim hanno proposto un metodo che, rispetto al precedente ha un'importante novità:

4) il metodo di classificazione TFCM (Thresholded Fuzzy C-Means). Tale metodo richiede che il valore del parametro ß sia scelto solamente quando per tutte le unità sono stati calcolati i valori definitivi del grado di appartenenza mediante il metodo delle k-medie sfocato senza alcuna variazione; a questo punto, detti μ_{ij} tali valori, si potrà scegliere ß in base all'analisi dei risultati ottenuti nonché tenendo conto che il massimo valore che si può assegnare a ß è dato da:

$$b_{\max} = \min\left\{ \min_i \max_j m_{ij}, \min_j \max_i m_{ij} \right\} \qquad (3.7)$$

Una volta scelto il valore di ß, tramite la (3.6) si procede alla normalizzazione ricavando i nuovi valori dei gradi di appartenenza.

I vantaggi di questo metodo rispetto al precedente sono due: anzitutto, sapendo quale è il valore massimo consentito per ß, si eliminano gli inconvenienti causati dalla scelta di un valore troppo alto; inoltre, poiché ß viene utilizzato solamente alla fine del procedimento, è possibile considerarne diversi valori per poter confrontare le classificazioni corrispondenti senza che ciò comporti di dover ripetere per ogni valore di ß l'intera procedura come invece accadeva con il metodo precedente. Questo, ovviamente, consente di risparmiare una grande quantità di tempo computazionale.

3.3.2. Il metodo delle k-medie semisfocato

I quattro metodi precedentemente esposti, pur se riescono a superare alcune problematiche del metodo delle k-medie sfocato, non risolvono appieno uno dei suoi maggiori difetti che consiste nel fatto che, una volta scelto il numero c di cluster nel quale le unità devono essere classificate, il procedimento per mezzo del quale per ciascuna unità viene determinato il grado di appartenenza μ_{ik} ad ognuno dei cluster non contempla il caso in cui un'unità non appartenga affatto ad uno o più cluster oppure, al limite, appartenga solamente ad uno di essi.

Per ogni unità, infatti, viene determinato un grado di appartenenza ad ognuno dei c cluster maggiore di zero non ammettendo in tal modo che il legame tra un'unità ed uno o più cluster possa essere nullo.

L'unica eccezione è rappresentata dal caso (molto raro, in verità) in cui un'unità coincide con il centro di un cluster, poiché in questo caso essa viene attribuita interamente a tale cluster ed il grado di appartenenza per tutti gli altri cluster viene posto uguale a zero (per rispettare il vincolo secondo il quale per ogni unità la somma dei gradi di appartenenza ai vari cluster deve essere uguale ad uno).

Per eliminare o, comunque, ridurre questo inconveniente, si è pensato quindi di apportare una nuova modifica al metodo delle k-medie sfocato che risponde all'esigenza di tenere in considerazione il fatto che, oltre alle unità coincidenti con il centro del cluster, anche alcune altre unità potrebbero possedere i requisiti necessari per essere classificate come totalmente appartenenti ad un solo cluster.

Nel metodo delle k-medie semisfocato (Iacovacci, 1997) si propone dunque di assegnare totalmente l'unità i-esima al cluster k-esimo se, detta d_{ik} la loro distanza, si ha $d_{ik} < (1 / \alpha) D$ dove α è un parametro (>1) determinato a priori, e D indica la distanza tra il centro del k-esimo cluster e il centro del cluster ad esso più vicino.

Questa formulazione risponde all'esigenza intuitiva di classificare come totalmente appartenente ad un cluster qualunque unità che abbia, rispetto al centro del cluster stesso, non solo una distanza ragionevolmente piccola, ma che inoltre sia abbastanza lontana da tutti gli altri cluster. Infatti, nel caso in cui l'unità i-esima fosse molto vicina al centro del cluster k-esimo, il quale è a sua volta poco distante dal centro del cluster p-esimo, classificare detta unità come totalmente appartenente al cluster k-esimo sarebbe un errore essendo in tal caso più giusto attribuire l'unità i-esima parte all'uno e parte all'altro cluster.

L'algoritmo che descrive il metodo delle k-medie semisfocato è il seguente:

Passo 1: si sceglie il valore del parametro m e quello di α e si ripartiscono le unità da classificare in c gruppi (con c scelto a priori), ottenendo la matrice iniziale $U^{(0)} = [\mu_{ik}]$, i=1,..,n; k=1,..,c (dove con l'esponente si indica il numero di iterazione).

Passo 2: si calcolano i centri dei gruppi usando la formula:

$$V_k^{(0)} = \frac{\sum_{i=1}^{n} (m_{ik})^m x_i}{\sum_{i=1}^{n} (m_{ik})^m} \qquad (3.8)$$

Passo 3: si calcola la matrice $D^{(0)}_1 = [d_{ik}]$, (con i=1,..,n; k=1,..,c) delle distanze delle unità dai centri dei cluster e la matrice $D_2^{(0)} = [d_{kk'}]$, (con k', k=1,..,c ; k' ≠ k) delle distanze tra i centri dei cluster.

Passo 4: si calcola la nuova matrice $U^{(1)}$ secondo le seguenti regole:

a) se per qualche gruppo r si ha che $d_{ir} = 0$, si pone $\mu_{ir} = 1$ e $\mu_{ik} = 0$ per tutti i $k \neq r$;
b) se $d_{ir} < (1/\alpha)\, d_{rk}$ per qualche r, con $k \neq r$, si pone $\mu_{ir} = 1$ e $\mu_{ik} = 0$ per tutti i $k \neq r$;
c) se nessuna delle due precedenti condizioni è soddisfatta allora si utilizza la seguente formula:

$$m_{ik} = \frac{1}{\sum_{j=1}^{c} \left[\frac{d_{ik}}{d_{jk}} \right]^{2/(m-1)}} \qquad (3.9)$$

Passo 5: si calcola la differenza $| U^{(1)} - U^{(0)} |$ e, se tale differenza risulta essere più piccola di un valore δ scelto a priori ed opportunamente piccolo, si ferma la procedura, altrimenti si torna al passo 2.

Una volta ottenuta la classificazione finale, è da notare che, qualora si desideri eliminare la presenza di valori molto bassi del grado di appartenenza, e rendere così più chiara la classificazione, può essere utilizzato anche il metodo TFCM (Thresholded Fuzzy C-Means) presentato nel paragrafo precedente con il quale, se $\mu_{ik} < ß$ (dove ß è un valore scelto a priori), si pone $\mu_{ik} = 0$. Ciò può essere fatto semplicemente aggiungendo il seguente:

Passo 6: se $\mu_{ik} < ß$, allora si pone $\mu_{ik} = 0$ e, per rispettare il vincolo secondo il quale per ogni unità la somma dei valori del grado di appartenenza deve essere 1, si opera poi la normalizzazione (3.6).

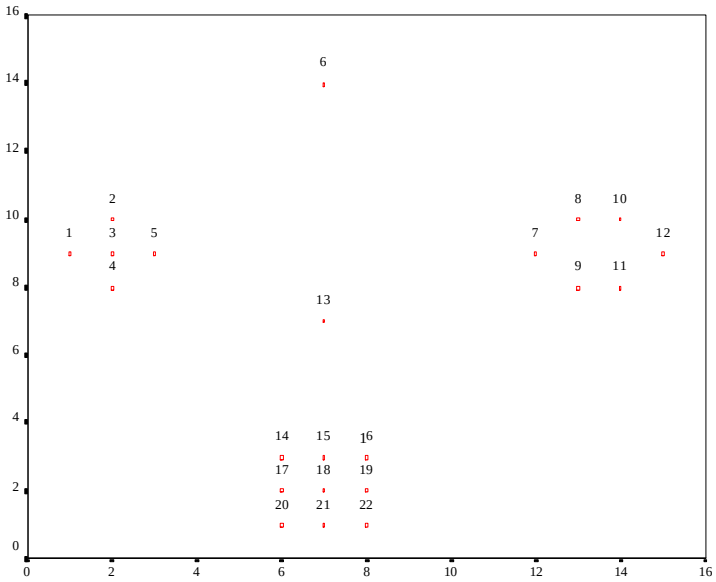
Per scegliere il valore del parametro α (che in genere deve essere superiore ad 1), sarà utile anche in questo caso utilizzare l'indice di sfocatura I proposto precedentemente.

La bontà della classificazione ottenuta utilizzando questo metodo rispetto a quella che si ottiene con il metodo sfocato, può essere brevemente descritta proponendo il seguente esempio.

3.3.3. Un esempio di applicazione del metodo delle k-medie sfocato e di quello semisfocato

Sulle 22 unità rappresentate in figura 3.1. è stato applicato sia il metodo delle k-medie sfocato che il metodo delle k-medie semisfocato utilizzando, in ambedue i casi, la norma euclidea, prendendo per il parametro δ il valore di 0.001, ed infine, poiché i cluster presenti sono visibilmente tre, assumendo questo valore per il parametro c.

Figura 3.1. Rappresentazione grafica delle unità considerate.



I risultati delle diverse applicazioni del metodo delle k-medie sfocato a seconda di alcuni valori di m sono dati nelle tabelle 3.1., 3.2., 3.3., mentre nella tabella 3.4. è riportato il risultato dell'applicazione del metodo delle k-medie semisfocato con i valori di m e di α fissati in base alle indicazioni fornite precedentemente dall'indice I.

A proposito dei risultati ottenuti, si noti inoltre che per ognuna delle singole prove la partizione iniziale è stata cambiata più volte, ottenendo sempre lo stesso risultato.

Tabella 3.1. ***Classificazione ottenuta con m=1.***

	grado di appartenenza		
Unità	Cluster 1	Cluster 2	Cluster 3
1	1.00	0.00	0.00
2	1.00	0.00	0.00

Tabella 3.2. ***Classificazione ottenuta con m=2.7.***

	grado di appartenenza		
Unità	Cluster 1	Cluster 2	Cluster 3
1	0.86	0.06	0.08
2	0.89	0.05	0.06

3	1.00	0.00	0.00
4	1.00	0.00	0.00
5	1.00	0.00	0.00
6	1.00	0.00	0.00
7	0.00	1.00	0.00
8	0.00	1.00	0.00
9	0.00	1.00	0.00
10	0.00	1.00	0.00
11	0.00	1.00	0.00
12	0.00	1.00	0.00
13	1.00	0.00	0.00
14	0.00	0.00	1.00
15	0.00	0.00	1.00
16	0.00	0.00	1.00
17	0.00	0.00	1.00
18	0.00	0.00	1.00
19	0.00	0.00	1.00
20	0.00	0.00	1.00
21	0.00	0.00	1.00
22	0.00	0.00	1.00
I = 0			

3	0.97	0.01	0.02
4	0.85	0.05	0.09
5	0.90	0.04	0.06
6	0.43	0.35	0.22
7	0.08	0.82	0.10
8	0.05	0.89	0.06
9	0.06	0.86	0.08
10	0.06	0.88	0.06
11	0.06	0.86	0.08
12	0.07	0.83	0.10
13	0.36	0.26	0.38
14	0.11	0.08	0.81
15	0.07	0.06	0.87
16	0.09	0.09	0.82
17	0.08	0.06	0.86
18	0.00	0.00	1.00
19	0.06	0.07	0.87
20	0.10	0.08	0.82
21	0.07	0.06	0.87
22	0.08	0.09	0.83
I = 0.411			

Tabella 3.3. ***Classificazione ottenuta con m=20.***

	grado di appartenenza		
Unità	Cluster 1	Cluster 2	Cluster 3
1	0.39	0.30	0.31
2	0.39	0.30	0.31
3	0.62	0.19	0.19
4	0.38	0.30	0.32
5	0.39	0.30	0.31
6	0.34	0.34	0.32
7	0.30	0.38	0.32
8	0.30	0.39	0.31
9	0.30	0.39	0.31
10	0.31	0.38	0.31
11	0.30	0.38	0.32
12	0.31	0.38	0.31
13	0.33	0.33	0.34
14	0.32	0.31	0.37
15	0.31	0.31	0.38
16	0.31	0.31	0.38
17	0.31	0.30	0.39
18	0.17	0.17	0.66
19	0.31	0.31	0.38
20	0.31	0.31	0.38
21	0.31	0.30	0.39
22	0.31	0.31	0.38
I = 0.976			

Tabella 3.4. ***Classificazione ottenuta con m=2.7 e a=3.***

	grado di appartenenza		
Unità	Cluster 1	Cluster 2	Cluster 3
1	1.00	0.00	0.00
2	1.00	0.00	0.00
3	1.00	0.00	0.00
4	1.00	0.00	0.00
5	1.00	0.00	0.00
6	0.43	0.35	0.22
7	0.00	1.00	0.00
8	0.00	1.00	0.00
9	0.00	1.00	0.00
10	0.00	1.00	0.00
11	0.00	1.00	0.00
12	0.00	1.00	0.00
13	0.36	0.26	0.38
14	0.00	0.00	1.00
15	0.00	0.00	1.00
16	0.00	0.00	1.00
17	0.00	0.00	1.00
18	0.00	0.00	1.00
19	0.00	0.00	1.00
20	0.00	0.00	1.00
21	0.00	0.00	1.00
22	0.00	0.00	1.00
I = 0.089			

Osservando i dati riportati nelle tabelle, si possono mettere in evidenza non solo le differenze tra le classificazioni sfocate ottenute a seconda del metodo utilizzato, ma, inoltre, è possibile osservare i vantaggi offerti dalla classificazione sfocata

rispetto a quella classica e confrontare le diverse classificazioni ottenute utilizzando alcuni valori di m.

Dall'esame della figura 3.1. risulta chiaro che le unità 6 e 13 non possono essere attribuite in alcun modo (a differenza delle altre unità) ad uno soltanto dei cluster essendo poste in una posizione intermedia rispetto ai cluster presenti (si noti in particolare l'unità 13).

Applicando il metodo delle k-medie classico (cioè scegliendo m=1) si ottengono i risultati posti nella tabella 3.1., dalla quale si può notare che le unità 6 e 13 vengono attribuite al cluster 1 del quale fanno parte le unità 1,2,3,4,5 ottenendo così una classificazione non rispondente alla realtà.

Aumentando il valore di m si ottiene una classificazione sempre più sfocata come l'indice I ben dimostra (si veda la classificazione ottenuta con m=20) e, per m=2.7, si ha un valore dell'indice I uguale a 0.411 al quale corrisponde la classificazione che meglio sembra rappresentare la situazione poiché le unità 6 e 13 vengono assegnate ai tre cluster che le attraggono in misura differente a seconda della loro vicinanza o meno da questi, mentre le restanti unità continuano ad avere un forte grado di appartenenza ai cluster ai cui centri sono molto vicine.

Il risultato ottenuto utilizzando il metodo delle k-medie sfocato con m=2,7 però, seppur migliore di quello fornito dal metodo classico non appare essere pienamente soddisfacente poiché, come si può vedere, nessuna delle unità che pure chiaramente appartengono ad un solo cluster ottiene il massimo valore del grado di appartenenza a meno che non si utilizzi un valore di m molto vicino ad 1 che però fornisce una classificazione troppo poco sfocata che male mette in risalto la situazione particolare delle unità 6 e 13.

Questo tipo di inconveniente non è invece presente nel metodo delle k-medie semisfocato il quale, utilizzando lo stesso valore m=2,7 e α=3, fornisce quello che sembra essere il risultato più rispondente alla reale classificazione delle unità poiché limita la sfocatura ai soli casi in cui essa è veramente esistente (si veda tab. 3.4.).

In conclusione, i vantaggi offerti da questo nuovo metodo consistono nel fatto che la sfocatura viene ad essere attribuita solamente alle unità la cui posizione è realmente incerta rispetto ai vari cluster (vedi nell'esempio le unità 6 e 13) mentre per le altre unità (a differenza del metodo delle k-medie sfocato) la sfocatura può anche essere eliminata.

Nel paragrafo seguente viene presentata un’applicazione del metodo delle k-medie semisfocato ad un caso reale.

3.3.4. Un’applicazione del metodo delle k-medie semisfocato alla classificazione dei comuni secondo il grado di urbanità e ruralità

La classificazione dei comuni secondo il grado di urbanità e ruralità che viene qui presentata rappresenta un miglioramento di quella precedentemente ottenuta utilizzando il metodo delle k-medie sfocato (Iacovacci, 1995). Ambedue le classificazioni, in ogni caso, si propongono di fornire un risultato più aderente alla realtà rispetto alla classificazione dell’ISTAT (1986) ricavata con il metodo delle k-medie classico.

In particolare nella classificazione ISTAT ogni comune viene assegnato ad una sola classe di appartenenza che ne identifica il grado di urbanità (o ruralità) posse-

duto prendendo in considerazione quattro diverse tipologie di comuni che possono essere così descritte:

1) **Comuni rurali:** essi sono caratterizzati dall'essere zone nelle quali sono ancora prevalenti strutture elementari, sia dal punto di vista delle attività lavorative (dove ha prevalenza assoluta o relativa l'attività agricola), sia per quel che riguarda i modi di vita che sono caratterizzati dall'essere a diretto contatto con la natura e appaiono essere molto omogenei riguardo alle tradizioni, ai costumi, alle opinioni, al linguaggio, ecc.
2) **Comuni semirurali:** si differenziano dai comuni rurali in quanto, pur possedendo ancora molte delle loro caratteristiche peculiari, è possibile individuare una certa differenziazione nei modi di vita che risentono in misura più accentuata delle trasformazioni avvenute nella società moderna.
3) **Comuni semiurbani:** sono quei territori nei quali il processo di trasformazione è più avanzato; in essi cioè coesistono, ancor più che nei comuni semirurali, strutture e organizzazioni proprie dell'ambiente cittadino mediate dal bagaglio culturale e comportamentale del mondo rurale.
4) **Comuni urbani:** in questo tipo di comuni è stata cancellata ogni traccia (in maniera più o meno profonda) dei modi di vita rurali e, per questo motivo, i comuni di questo tipo appaiono essere nettamente più caratterizzati dei comuni semiurbani.

Dalle definizioni date, risulta abbastanza evidente che solamente nei comuni rurali si può ipotizzare che in tutto il territorio esista una certa omogeneità nei modi di vita, giacché a proposito dei comuni urbani occorre sottolineare che, nel loro ambito, coesistono le città con milioni di abitanti e i centri di minor dimensione, le città caratterizzate da attività prevalentemente industriale e quelle dedite soprattutto all'attività commerciale e di servizio, e così via, mentre nei restanti due casi intermedi la situazione è piuttosto eterogenea, poiché l'elemento rurale è sempre presente talvolta in misura minore (comuni semiurbani), talaltra in misura prevalente (comuni semirurali).

Per questo motivo, quindi, l'adottare una classificazione di tipo classico non risulta molto indicato per il fenomeno considerato poiché solamente in casi singolari si può supporre che nell'intero territorio comunale esista un unico modo di vita corrispondente ad un certo grado di urbanità. Nella grande maggioranza dei comuni, infatti, coesistono diversi aspetti propri dei vari gradi di urbanità che vanno dall'urbano al rurale, perciò per avere una classificazione che rispecchi la reale situazione dei comuni, è necessario utilizzare una procedura che, per ogni comune, sia in grado di distinguere la componente urbana, semiurbana, semirurale e rurale.

Il tipo di classificazione che risponde a questa esigenza è la classificazione sfocata, ed in particolare, si è utilizzato il metodo delle k-medie semisfocato descritto in precedenza per classificare gli 8086 comuni italiani (dati del Censimento 1981) secondo il grado di urbanità e ruralità posseduto.

Per ottenere una classificazione sfocata che sia confrontabile con quella classica dell'ISTAT del 1986, allo scopo di poter così valutare appieno i vantaggi propri della nuova classificazione, sono state impiegate le stesse variabili e gli stessi dati

utilizzati dall'ISTAT(1.). Tali variabili, sulla cui scelta valgono le considerazioni svolte dall'ISTAT, sono:

v_1: densità (abitanti per km²);
v_2: percentuale della popolazione attiva in condizione professionale sul totale della popolazione in età da 14 anni in poi;
v_3: percentuale della popolazione attiva in agricoltura sul totale della popolazione attiva in condizione professionale;
v_4: percentuale delle donne attive nei settori extra agricoli sul totale della popolazione femminile in età da 14 anni in poi;
v_5: percentuale delle persone in possesso di laurea o titolo di studio di scuola media superiore sul totale della popolazione in età da 18 anni in poi;
v_6: percentuale degli occupati con luogo di lavoro situato all'esterno del comune sul totale degli occupati;
v_7: percentuale del totale degli "addetti" alle unità locali sulla popolazione in età da 14 anni in poi;
v_8: percentuale degli "addetti" alle unità locali del settore terziario (commercio escluso) sulla popolazione in età da 14 anni in poi;
v_9 : numero medio di componenti per famiglia;
v_{10}: percentuale delle abitazioni godute in proprietà sul totale delle abitazioni occupate;
v_{11}: percentuale delle abitazioni fornite di alcuni servizi all'interno dell'abitazione (acqua potabile ed acquedotto, gabinetto) sul totale delle abitazioni occupate;
v_{12}: percentuale delle utenze telefoniche (totale) sulla popolazione;
v_{13}: percentuale delle utenze telefoniche ("affari") sulla popolazione.

In generale, le vari fasi del procedimento di classificazione utilizzato possono essere così brevemente descritte:

a) standardizzazione dei dati in modo tale da avere media nulla e varianza unitaria;
b) applicazione sui dati standardizzati del metodo delle k-medie semisfocato assumendo c = 4, la classificazione dell'ISTAT come partizione iniziale $U^{(0)}$, $\delta = 0{,}01$ e come metrica quella euclidea;
c) ripetizione dell'intera procedura con diversi valori di m e di α al fine di scegliere il migliore;
d) fissati i valori m=1,3 e α=1,3 ed ottenuta la classificazione corrispondente, utilizzo del parametro β con valore 0.02 in modo da rendere nulli i valori μ_{ik} inferiori a tale valore;
e) ripetizione dell'intera procedura utilizzando differenti partizioni iniziali e valutazione di eventuali differenze nelle classificazioni ottenute in corrispondenza.

Una volta fissati i valori di m e di α e ottenuta la classificazione cercata, l'intera procedura è stata ripetuta utilizzando diverse partizioni iniziali per valutare, come già detto, se ed in quale misura la scelta della partizione iniziale fosse influente sulla classificazione finale. Le partizioni utilizzate, in particolare, sono state le seguenti:

a) partizione casuale (o random), dove l'assegnazione iniziale di ogni comune ad una delle quattro classi predisposte è stata fatta casualmente dall'elaboratore;

---

(1) A tal proposito, si ringraziano l'ISTAT per la gentile concessione dei dati e la società SARIN per aver autorizzato l'ISTAT a concederci di utilizzare i dati sull'utenza telefonica nel 1982.

b) partizione sistematica 1, nella quale gli 8086 comuni, ordinati secondo il codice ISTAT, sono stati divisi in quattro gruppi mediante i tre quartili, assegnando poi i comuni del primo gruppo alla prima classe, quelli del secondo gruppo alla seconda classe e così via;
c) partizione sistematica 2, dove il comune i è stato assegnato alla classe h (h = 1, 2, 3, 4) dove h è il resto della divisione di i per 4 (precisando che, quando il valore del resto è 0, il comune va assegnato alla classe 4).

Le tre classificazioni ottenute in corrispondenza delle tre diverse partizioni iniziali, sono risultate identiche tra loro e con quella ottenuta utilizzando come partizione iniziale la classificazione dell'ISTAT, sia per quanto riguarda i valori della funzione di appartenenza, sia per quanto riguarda il valore finale della funzione di ottimizzazione, lasciando ritenere così che questo valore corrisponda ad un valore di minimo globale e, dunque, di ottimo assoluto. Bisogna inoltre aggiungere che la stabilità della soluzione ottenuta fa concludere che anche il numero di cluster sia stato ben scelto. Le uniche differenze riscontrate a seconda del tipo di partizione iniziale adoperata, consistono nel differente numero di iterazioni occorrenti per ottenere la classificazione finale. Il numero delle iterazioni è riportato nella tabella seguente:

Tabella 3.5. ***N° di iterazioni per diverse partizioni iniziali.***

Partizione iniziale	Iterazioni
ISTAT	73
Casuale	116
Sistematica 1	117
Sistematica 2	96

Come si può notare dalla tabella, il minor numero di iterazioni si è avuto quando come partizione iniziale è stata utilizzata la classificazione dell'ISTAT, cioè una classificazione che, rispetto alle altre, è più vicina a quella finale. Allo scopo di ridurre i tempi di elaborazione, questo dato suggerisce quindi di utilizzare in generale come partizione iniziale non una classificazione qualunque, ma, quando essa sia disponibile, una classificazione ottenuta precedentemente utilizzando altri metodi di classificazione. Questo modo di procedere può essere inoltre giustificato dal fatto che il metodo delle k-medie semisfocato, così come quello sfocato, sembra essere piuttosto robusto rispetto alla scelta della partizione iniziale.

Dopo aver descritto la procedura impiegata ed averne illustrato alcuni aspetti particolari, verranno ora presentati i risultati ottenuti.

Tabella 3.6. ***Classificazione semisfocata dei comuni della provincia di Viterbo.***

Codice	Comune	1	2	3	4	T
56001	Acquapendente	0.045	0.056	0.596	0.303	3
56002	Arlena di Castro	0.000	0.021	0.303	0.676	4
56003	Bagnoregio	0.029	0.038	0.751	0.182	3
56004	Barbarano Romano	0.000	0.000	0.571	0.429	3
56005	Bassano Romano	0.000	0.000	1.000	0.000	3
56006	Bassano in Teverina	0.021	0.054	0.692	0.233	3
56007	Blera	0.000	0.000	0.901	0.099	3
56008	Bolsena	0.000	0.000	0.409	0.591	3
56009	Bomarzo	0.000	0.000	0.715	0.285	3
56010	Calcata	0.000	0.000	0.741	0.259	3
56011	Canepina	0.000	0.000	1.000	0.000	4
56012	Canino	0.000	0.000	0.205	0.795	4
56013	Capodimonte	0.000	0.020	0.782	0.198	3
56014	Capranica	0.000	0.000	1.000	0.000	3
56015	Caprarola	0.000	0.000	0.344	0.656	3
56016	Carbognano	0.000	0.000	0.124	0.876	4
56017	Castel Sant'elia	0.040	0.167	0.709	0.084	3
56018	Castiglione in Teverina	0.000	0.000	1.000	0.000	3
56019	Celleno	0.328	0.193	0.305	0.174	1
56020	Cellere	0.000	0.000	0.185	0.815	4
56021	Civita Castellana	0.569	0.252	0.146	0.033	1
56022	Civitella d'Agliano	0.000	0.000	0.551	0.449	3
56023	Corchiano	0.000	0.000	0.000	1.000	4
56024	Fabrica di Roma	0.000	0.167	0.716	0.117	3
56025	Faleria	0.000	0.000	0.508	0.492	3
56026	Farnese	0.000	0.000	0.156	0.844	4
56027	Gallese	0.027	0.147	0.718	0.108	3
56028	Gradoli	0.000	0.000	0.089	0.911	4
56029	Graffignano	0.000	0.000	1.000	0.000	3
56030	Grotte di Castro	0.000	0.000	0.000	1.000	4
56031	Ischia di Castro	0.000	0.000	0.053	0.947	4
56032	Latera	0.000	0.000	0.210	0.790	4
56033	Lubriano	0.000	0.000	0.535	0.465	3
56034	Marta	0.000	0.000	0.462	0.538	3
56035	Montalto di Castro	0.022	0.034	0.276	0.668	4
56036	Montefiascone	0.000	0.000	1.000	0.000	3
56037	Monte Romano	0.000	0.036	0.556	0.408	3
56038	Monterosi	0.000	0.062	0.444	0.494	4
56039	Nepi	0.000	0.000	1.000	0.000	3
56040	Onano	0.000	0.000	0.091	0.909	4
56041	Oriolo Romano	0.000	0.074	0.769	0.157	3
56042	Orte	0.830	0.052	0.091	0.027	1
56043	Piansano	0.000	0.000	0.148	0.852	4
56044	Proceno	0.000	0.000	0.271	0.729	3
56045	Ronciglione	0.489	0.082	0.369	0.060	1
56046	Villa S. Giovanni Tuscia	0.000	0.039	0.747	0.213	3
56047	San Lorenzo Nuovo	0.000	0.000	0.249	0.751	4
56048	Soriano nel Cimino	0.000	0.000	1.000	0.000	3
56049	Sutri	0.000	0.023	0.863	0.114	3
56050	Tarquinia	0.585	0.087	0.228	0.100	1
56051	Tessennano	0.000	0.025	0.316	0.659	4
56052	Tuscania	0.034	0.048	0.554	0.365	3
56053	Valentano	0.000	0.000	0.645	0.355	3
56054	Vallerano	0.000	0.000	0.000	1.000	3
56055	Vasanello	0.000	0.000	0.477	0.523	3
56056	Vejano	0.000	0.046	0.597	0.357	3
56057	Vetralla	0.000	0.000	1.000	0.000	3
56058	Vignanello	0.000	0.000	0.000	1.000	4
56059	Viterbo	1.000	0.000	0.000	0.000	1
56060	Vitorchiano	0.000	0.000	1.000	0.000	3

Essendo impossibile, per ragioni di spazio, riportare la classificazione completa di ogni singolo comune, qui, a titolo di esempio, riportiamo la classificazione sfocata dei comuni della provincia di Viterbo nella quale sono presenti alcuni comuni che ben si prestano ad evidenziare i vantaggi forniti dalla classificazione sfocata. Nella tabella 3.6. sono specificati, per ogni comune, il codice ISTAT, il valore del grado di appartenenza per ognuna delle quattro classi (specificando che la classe 1 corrisponde ai comuni urbani, la 2 ai comuni semiurbani, la 3 ai comuni semirurali e la 4 ai comuni rurali), e la classe, indicata nella colonna contrassegnata dalla lettera T, alla quale il comune era stato assegnato nella classificazione ISTAT del 1986.

Da tale tabella si può vedere che, nella classificazione ISTAT, vi sono sei comuni (Celleno, Civita Castellana, Orte, Ronciglione, Tarquinia e Viterbo) che vengono tutti indistintamente attribuiti alla classe 1 dei comuni urbani; con tale classificazione dunque, un comune certamente urbano come Viterbo viene assimilato ad altri comuni che, pur presentando anch'essi dei forti connotati urbani, senza dubbio però lo sono in maniera minore.

Questo aspetto è invece ben rilevato dalla classificazione sfocata la quale mentre definisce completamente urbano il comune di Viterbo, definisce gli altri 5 comuni come appartenenti alla classe 1 con gradi decisamente inferiori (la percentuale più alta, ad esempio, è quella di Orte con l'83 % seguita da Tarquinia con il 58.5%) poiché tutti presentano anche forti caratteristiche semiurbane (Civita Castellana 25.2% e Celleno 19.3%) o, addirittura, semirurali (Ronciglione 36.9%, Celleno 30.5%, Tarquinia 22.8%).

Lo stesso discorso può essere ripetuto con evidenza ancora maggiore per gli altri comuni che la classificazione ISTAT definisce come semirurali o rurali: si noti infatti come praticamente tutti i comuni appartenenti alla classe 3 secondo la classificazione dell'ISTAT, nella classificazione sfocata presentino invece un forte o, comunque, rilevante grado di appartenenza anche alla classe 4 (è il caso, ad esempio, di Bolsena, Caprarola, Marta, Proceno e Vasanello dove addirittura è più forte il grado di appartenenza alla classe 4, o di Barbarano Romano, Civitella d'Agliano, Faleria, Lubriano, Tuscania, ecc. per i quali il legame alla classe 4 è superiore al 30%); tra i comuni definiti dalla classificazione ISTAT come comuni rurali, si veda, ad esempio, Montalto di Castro, Monterosi, San Lorenzo Nuovo, ecc. i quali ora ottengono un grado di appartenenza superiore al 20% anche alla classe 3.

Infine, come ulteriore considerazione sulla bontà della classificazione ottenuta, si noti come, nella grande maggioranza dei casi, ogni comune risulti avere un forte grado di appartenenza in prevalenza alla classe alla quale era stato assegnato dall'ISTAT e, per la parte rimanente, a delle classi che, in genere, sono ad essa contigue, mentre, tra i comuni assegnati totalmente ad una sola classe, solamente due comuni, Canepina e Vallerano, risultano assegnati a due classi che non coincidono con quelle dell'ISTAT. È importante sottolineare inoltre che nessun comune ottiene contemporaneamente un forte grado di appartenenza alle due classi estreme di urbanità evitando così dei possibili risultati di dubbio significato.

Per concludere il commento dei risultati, riportiamo ora qualche dato generale relativo alle altre province.

Degli 8086 comuni classificati, 145 sono risultati completamente urbani, 384 semiurbani, 191 semirurali e 594 rurali mentre i restanti comuni sono risultati appartenere a due o più classi con diversi gradi di appartenenza. È inoltre da sottoli-

neare che dei 145 comuni urbani ben 25 sono dei capoluoghi di provincia (tra i quali vi sono, per esempio, Vercelli, Venezia, Parma, Siena, ecc.) situati in gran parte in Piemonte, Veneto, Emilia Romagna e Toscana, mentre altri capoluoghi di provincia (anche questi prevalentemente del Nord e del Centro), come per esempio Aosta, Bolzano, Verona, Pisa, Perugia, Frosinone, ecc. sono legati solamente alle classi 1 e 2 con un grado di appartenenza anche superiore al 95% per la classe 1. Tra gli altri comuni che appartengono solamente alla classe 1, sono da segnalare fra gli altri quelli di Saint-Vincent (Ao), Riva del Garda (Tn), Rimini (Fo), Senigallia (An), ecc. mentre diversi comuni tra i quali Anzio (Roma) e Civitavecchia (Roma) appartengono solamente a due classi (in questo caso la 1 e la 2).

Il valore assunto dall'indice I è 0.432.

In conclusione, l'applicazione del metodo delle k-medie semisfocato ai comuni italiani secondo il grado di urbanità ha fornito dei risultati che evidenziano, in modo concreto, quali siano i vantaggi offerti dalla classificazione sfocata. Essi possono essere riassunti brevemente col dire che, mentre con la classificazione sfocata è possibile stabilire per ogni comune in quale proporzione sono presenti le differenti componenti dell'urbanità che spesso si trovano a convivere nell'ambito dello stesso comune, individuando il grado di appartenenza del comune stesso ad ogni classe (le quali possono essere distinte nella classe dei comuni rurali, semirurali, semiurbani e urbani), ciò invece non è possibile con la classificazione classica la quale, attribuendo interamente ogni comune ad una sola di queste classi, di fatto è in grado di individuare solamente quale è la componente prevalente nel comune, fornendo dunque delle indicazioni certamente più approssimate.

La classificazione dei comuni secondo il grado di urbanità e ruralità ottenuta con il metodo delle k-medie semisfocato, inoltre, è un esempio di come esso sembra essere migliore del metodo sfocato poiché nella classificazione ottenuta, pur rimanendo sfocata e risultando in gran parte simile alla classificazione ottenuta con il metodo sfocato, sono eliminati quei casi in cui la sfocatura era prodotta "artificialmente" dal metodo stesso.

3.4. Altri metodi non gerarchici di classificazione sfocata

Oltre al metodo delle k-medie sfocato ed ai suoi perfezionamenti presentati nei paragrafi precedenti, esiste in letteratura un grande numero di metodi di classificazione sfocata di tipo non gerarchico. Poiché descriverli tutti esula dalle finalità del presente lavoro, nei prossimi paragrafi saranno portati ad esempio solamente due di essi.

3.4.1. Il metodo FUNNY

FUNNY è il nome di un programma per personal computer il quale produce delle classificazioni sfocate utilizzando qualunque tipo di dati (siano essi numerici o misure di dissimilarità).

Il procedimento attraverso il quale si perviene alla classificazione delle unità è molto simile a quello utilizzato nel metodo delle k-medie sfocato: anche in questo caso, infatti, fissato il numero di gruppi nei quali si desidera suddividere le unità, si fornisce una partizione iniziale delle unità stesse assegnando a priori i valori della

funzione di appartenenza; i valori finali si ottengono poi eseguendo una procedura di ottimizzazione di questi valori attraverso l'uso di una funzione obiettivo.

La differenza maggiore tra il metodo FUNNY (Kaufmann,1990) ed il metodo delle k-medie sfocato consiste proprio nella funzione obiettivo che nel metodo FUNNY assume la seguente forma:

$$C = \sum_{k=1}^{c}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}(m_{ik})^2(m_{jk})^2(d_{ij}) / 2\sum_{j=1}^{n}(m_{jk})^2\right] \qquad (3.10)$$

dove (d_{ij}) rappresenta la distanza (o la dissimilarità) tra l'unità i-esima e l'unità j-esima e μ_{ik} esprime il grado di appartenenza dell'unità i al cluster k.

La soluzione di minimo per la funzione (3.10) si trova applicando il metodo dei moltiplicatori di Lagrange sotto le condizioni di Kuhn e Tucker alla funzione (3.10) sotto i vincoli :

i) $\mu_{ik} \geq 0$ i=1,...,n k=1,...,c

ii) $\sum_{k=1}^{c} m_{ik} = 1$ i=1,...,n

Tale soluzione è in genere una soluzione di ottimo locale e si ottiene per mezzo di una procedura di ottimizzazione che qui, per brevità di esposizione, non esporremo.

Come detto, la differenza tra il metodo FUNNY e il metodo delle k-medie sfocato risiede solamente nella differente funzione obiettivo adottata, ed in particolare nel fatto che il metodo delle k-medie sfocato considera il quadrato della distanza, mentre il metodo FUNNY la utilizza con l'esponente uguale ad 1; inoltre, mentre nel metodo delle k-medie sfocato il parametro m può variare (m > 1), nel metodo FUNNY esso è fisso e posto uguale a 2 poiché, secondo l'autore, è con tale valore che pare si ottengano i risultati migliori (notiamo comunque che, variando l'esponente, si ottengono delle classificazioni più o meno sfocate proprio come col metodo delle k-medie sfocato).

Per avere un confronto tra i 2 metodi, nella tabella 3.7. sono riportati i risultati ottenuti applicando il metodo delle k-medie sfocato (con m=2) ed il metodo FUNNY ai dati della figura 3.1.

Tabella 3.7. ***Classificazioni ottenute con i metodi delle k-medie sfocato e FUZZY.***

Unità	Metodo delle k-medie sfocato			Metodo FUZZY		
	Grado di appartenenza					
	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3
1	0.96	0.01	0.03	0.87	0.06	0.07
2	0.98	0.01	0.01	0.88	0.05	0.07
3	0.98	0.01	0.01	0.93	0.03	0.04
4	0.96	0.01	0.03	0.86	0.06	0.08
5	0.98	0.01	0.01	0.87	0.06	0.07
6	0.50	0.35	0.15	0.42	0.35	0.23
7	0.02	0.96	0.02	0.08	0.82	0.10
8	0.01	0.98	0.01	0.06	0.87	0.07
9	0.01	0.97	0.02	0.06	0.86	0.08
10	0.01	0.98	0.01	0.06	0.87	0.07
11	0.01	0.97	0.02	0.06	0.86	0.08
12	0.02	0.96	0.02	0.07	0.84	0.09
13	0.37	0.22	0.51	0.36	0.27	0.37
14	0.03	0.02	0.95	0.12	0.08	0.80
15	0.01	0.01	0.98	0.08	0.07	0.85
16	0.02	0.03	0.95	0.10	0.10	0.80
17	0.01	0.01	0.98	0.08	0.06	0.86
18	0.00	0.00	1.00	0.04	0.04	0.92
19	0.01	0.01	0.98	0.07	0.07	0.86
20	0.03	0.02	0.95	0.10	0.08	0.82
21	0.01	0.01	0.98	0.07	0.06	0.87
22	0.02	0.02	0.96	0.09	0.09	0.82

Dall'analisi dei risultati emerge chiaramente che, nella sostanza, entrambi i metodi producono (a parità di m) le stesse classificazioni; ricordando però che con il metodo delle k-medie sfocato si ha la possibilità di scegliere il valore migliore per m (che non sempre risulta essere m=2), quest'ultimo sembra essere preferibile.

3.4.2. Il metodo MND2

Un altro metodo di classificazione sfocata che presenta delle affinità con i metodi precedenti, è l'algoritmo MND2 di Roubens (Roubens, 1978) che si differenzia dagli altri metodi a causa della funzione obiettivo adottata che è la seguente:

$$R = \sum_{k=1}^{c} \sum_{i=1}^{n} \sum_{j=n}^{n} (m_{ik})^2 (m_{jk})^2 (d_{ij}) \qquad (3.11)$$

Sebbene tale funzione appare molto simile a quella adottata dal metodo FUNNY (la differenza maggiore consiste nel non considerare il denominatore), la sua minimizzazione però tende a distorcere i risultati, in particolare se il numero di unità da classificare è molto elevato. In tal caso, infatti, il metodo MND2 non sempre assicura una convergenza ottimale il che, abbinato all'altro difetto ora esposto, rende questo metodo decisamente meno preferibile rispetto a quelli considerati in precedenza.

3.5. Osservazioni

I metodi che sono stati esposti in questo capitolo presentano, come è stato visto, forti analogie tra di loro in quanto i procedimenti utilizzati per ottenere le classificazioni cercate sono tra loro molto simili. Poiché questi metodi sono tutti di tipo non gerarchico e alcuni di essi rappresentano una generalizzazione dei metodi classici, essi presentano in generale lo stesso tipo di problemi propri dei metodi non gerarchici di classificazione classica che consistono principalmente nella difficoltà di scelta delle condizioni iniziali.

Per quanto riguarda la scelta di alcuni parametri, nei paragrafi precedenti sono state proposte alcune soluzioni, ma diversi problemi restano di difficile soluzione come per esempio la scelta iniziale del numero c di gruppi la quale può essere fatta solamente dopo aver compiuto un'accurata analisi dei dati a disposizione e dopo aver ripetuto la procedura di classificazione per diversi valori di c in modo tale da poter valutare quale sembra essere il valore migliore.

Un altro inconveniente comune a questi metodi è poi quello relativo alla convergenza che non sempre risulta essere ottimale in quanto alle volte i risultati ottenuti corrispondono a soluzioni di minimo locale e non assoluto.

È infine da notare che, poiché il problema della convergenza è legato a quello della scelta della funzione obiettivo, quasi tutti i metodi di questo tipo utilizzano una funzione simile o riconducibile a quella utilizzata nel metodo delle k-medie sfocato giacché questo è quello che sembra convergere più rapidamente alla soluzione ottimale.

Nel complesso, comunque, i metodi precedentemente esaminati forniscono dei risultati che possono essere considerati molto soddisfacenti e, trattandosi di metodi tutti molto recenti, è facile prevedere che nell'immediato futuro verranno apportati ulteriori miglioramenti.

BIBLIOGRAFIA

Anderberg M. R. (1973), *Cluster analysis for applications*, New York, Academic Press.

Badaloni M. e Vinci E. (1983), "Osservazioni sugli indici di similarità", in *Metron*, 41(1-2), 113-133.

Ball G.H. e Hall D. J. (1967), "A clustering technique for summarizing multivariate data", in *Behavioral Sci*, 12, 153-155.

Bartko J., Strauss J. e Carpenter J. (1971), *An Evolution of Taxonometric Tecnique for Psychiatric Data*, Class. Soc. Bull., 2, 21-27.

Bellacicco A. (1977), "Clustering tume varyng data, in J.R. Barra e al., *Recent developpment in statisics*, North Holland, Amsterdam, 739-747.

Bellman e Prade (1970), "Decision-making in a fuzzy enviroment", in *Manage.Sci.*, 17, (4), B141-B164.

Bezdek J. C. (1981), *Pattern recognition with fuzzy objective function algorithms*, New York, Plenum.

Bezdek J. C. e Hathaway R. J. (1988), "Recent convergence results for the fuzzy c-means clustering algorithms" , in *J. of Classification*, 5, 237-247.

Bezdek J. C., Hathaway R. J. e Kim T. (1988), "Optimality tests for fixed points of the fuzzy c-means algorithm", in *Pattern Recognition*, 21(6), pp. 651-663.

Bezdek J. C., Hathaway R. J. e Windham M. P. (1991), "Numerical comparison of the RFCM and AP algorithms for clustering relational data", in *Pattern recognition*, 24(8), 783-791.

Bezdek J. C., Hathaway R. J. e Davenport J. W. (1989), "Relational duals of the c-means clustering algorithms", in *Pattern recognition*, 22(2), 205-212.

Bezdek. J.C. (1981). *Pattern Recognition with Fuzzy Objective Function Algoritm*, Plenum Press, New York.

Bezdek. J.C. (1991), "Numerical comparison of the RFCM and AP algorithms for clustering relational data", in *Pattern Recognition*, 24(8), 786-791.

Caselli G. (1988), *La mortalità per causa: caratteristiche attuali e recente evoluzione. 2° rapporto sulla situazione demografica italiana*, IRP, 4, 105-113.

Caselli G. (1993), *L'evolution à long term de la mortalité en Europe*, Proceedings of the European Conference, 2, Ined, Paris.

Caselli G., Cerbara L. e Leti G. (1993), "The geography of adult mortality: results from the fuzzy clumping method", in *Genus*, 49(1-2), 1-24.

Caselli G. e Egidi V. (1993), "Socio-economic development and regional differences in mortality in Europe", *Les comportements démographiques en Europe:*

facteurs de differenciation régional, 18-19 Febbraio, Université Libre de Bruxelles, Bruxelles.

Cerbara L. (1992), "The fuzzy climping method", in *Metron*, 50(3-4), 61-89.

Cerbara L. (1997), "Hierarchical fuzzy clustering: an example of spatio-temporal analysis", in *Book of Short Papers -Classification and Data Analysis*, Università degli Studi G. d'Annunzio di Pescara.

Deloche R. (1975), "Théorie des sous-ensamble flous et classification en analyse économique spatiale, *Document de Travail de l'I.M.E.*, 11.

Di Ciacco A. (1990), "Analisi simultanea dei caratteri qualitativi e quantitativi attraverso la parametrizzazione dei dati", in *Metron*, 48(1-4), 333-364.

Diday E. (1980), "Orders and Overlapping Clusters by Piramids, in J. De Leew e al., *Multidimentional Data Analysis*, DWSO Press, Leiden.

Doubois D. e Prade H. (1980), *Fuzzy Set and System: Theory and Application*, Accademic Press, New York.

Duda R. e Hart P. (1973), *Pattern classification and scene analysis*, New York, Wiley.

Dunn J. C. (1974), "Some recent investigations of a new fuzzy partitioning algorithm and its application to pattern classification problems", in *J. Cybernetic*, 4, 1-15.

Dunn J. C. (1974), "Well separated clusters and optimal fuzzy partitions", in *J. Cybernetic*, 4, 95-104.

Dunn J. C.(1974), "A fuzzy relative of the ISODATA process and its use in detecting compact, well separated clusters", in *J. Cybernetic*, 3, 32-57.

Everitt B. S. (1980), *Cluster analysis*, London, Heineman Educational Books.

Fowlkes E. B., Gnanadesikan R. e Kettenring J. R. (1988), "Variable selection in clustering", in *J. of Classification*, 5, 205-228.

Fustier B. (1975), "L'attraction des points de vente dans des espaces précis et imprécis", in *Document de Travail de l'I.M.E.*, 10.

Fustier B. (1980), "Contribution à l'étude d'un caractére statistique flou", in *Document de Travail de l'I.M.E.*, 38.

Giles R. (1988). The concept of grade of membership, in *Fuzzy Sets and Systems*, 25, 297-323.

Gordon A.D. (1981), *Classification - Monographs on Applied Probability and Statistics*, Chapman and Hall, London-New York.

Gordon A.D. e Vichi M. (1997), "Partitions of Partitions", sottoposto per la pubblicazione.

Hartigan J.(1975), *Clustering algorithms*, New York, Wiley.

Iacovacci G. (1995), “Sull’utilizzo del metodo delle c-medie sfocato per la classificazione dei comuni italiani a seconda del grado di urbanità e ruralità”, in *Statistica Applicata*, 7(1), Rocco Curto Editore.

Iacovacci G. (1997), “A semi-fuzzy c-means algorithm for soft clustering”, in *Book of Short Papers -Classification and Data Analysis*, Università degli Studi G. d’Annunzio di Pescara.

Kamel M.S. e Selim S.Z. (1991), “A thresholded fuzzy c-means algorithm for semi-fuzzy clustering”, in *Pattern Recognition*, 24(9), 825-883.

Kaufman L. and Rousseeuw P. J. (1990), *Finding groups in data*, Wiley, New York

Kaufmann A. (1975), *Introduction to the theory of fuzzy subsets*, 1, Academic Press, New York.

Kaufmann A. (1977), *Introduction à la théorie del sous-ensamble flous. 1.Eléments théoriques de base*, MASSON.

Kaufmann A.e Rousseeuw P. (1990), *Finding Groups in Data. An Introdution to Cluster Analysis*, Wiley, New York.

Leti G. (1979), *Distanze ed indici statistici*, La Goliardica editrice, Roma.

Leung Y. (1986), *Spatial analysis and planning under imprecision*, North-Holland, Amsterdam.

MacQueen J. (1967), “Some methods for classification and analysis of multivariate observations”, in *Proc. 5th Berkeley Symp. on Math. Stat and Prob*, 281-297.

Milioli M.A. (1994), “Confronto fra partizioni sfocate nell’analisi di dati territoriali” in *Atti della XXXVII Riunione Scientifica (SIS) San Remo 6-8 aprile*, 2.

Milligan, G. W. (1980), “An examination of the effect of six types of error perturbation on fifteen clustering algorithms”, in *Psychometrika*, 45, 325-342.

Milligan, G. W. e Cooper M. C. (1985), “The effect of error on determining the number of clusters in a data set”, in *Psychometrika*, 50, 159-179.

Milligan, G. W. e Cooper M. C. (1988), “A study of standardization of variables in cluster analysis”, in *J. of Classification*, 5, 181-204.

Mirkin (1990), “A sequential fitting procedure for linear data analysis”, in *Journal of Classification*, 167-195.

Ponsard C. (1985), “Fuzzy data analysis in a spatial context”, in. *Measuring the unmensurable*, P. Nijkamp et al., Martinus Nijoff Publishers, Dordrecht.

Ricolfi L. (1992), *HELGA - Nuovi principi di analisi dei gruppi*, FrancoAngeli, Milano.

Rizzi A. (1985), *Analisi dei dati. Applicazioni dell'informatica alla statistica*, Studi Superiori NIS.

Rolland-May (1985), “Fuzzy geografical space: algorithms of fuzzy and application to fuzzy regionalization”, *Sistemi Urbani* , 3.

Roubens M. (1978), “Pattern classification problems and fuzzy sets”, in *J. Fuzzy sets*, 1, 239-253.

Roubens M. (1982), “Fuzzy clustering algorithms and their cluster validity”, in European Journal of Operational Research, 10, 294-301.

Selim S. Z. e Ismail M. A. (1984), “Soft clustering of multidimensional data: a semi-fuzzy approach”, in *Pattern Recognition*, 17(5), 559-568.

Selim S. Z. e Kamel M. S. (1991), “A thresholded fuzzy c-means algorithm for semi-fuzzy clustering”, in *Pattern Recognition*, 24(9), 825-883.

Spath H. (1980), *Cluster analysis algorithms*, Chichester, E. Horwood.

Tran Qui P. (1978), *Les régions économiques floues: application au cas de la France, Colletion de l'I.M.E.*, 16, Librairie de l'Université, Dijon.

Wang, Hall e Subaryono (1990), “Fuzzy information representation and processing in conventional GIS software: data base design and application”, in *Journal Geographic Information System*, 261-283.

Zaddeh L.A. (1977), “Fuzzy set and their application to pattern classification and clustering”, in *Classification and Clustering*, J. Van Ryzin, Accademic Press, New York, 251-299.

Zadeh L. A. (1965), “Fuzzy sets”, *Inf. Control*, 8, 338-353.

Zani S. (1987), “Comparing Partitions of Geographical Areas”, *Contributed Papers 46h Session of ISI*, Tokyo, 499-500.

Zani S. (1988), Un metodo di classificazione "sfocata", in G. Diana, C. Provasi e R. Vedaldi, *Metodi statistici per la tecnologia e l'analisi dei dati multidimensionali*, Università degli Studi di Padova, 281-288.

Zani S. (1989), “Classificazioni "sfocate" di unità territoriali: un'applicazione alle regioni italiane”, in *Atti delle Giornate di studio “Analisi Statistica di Dati Territoriali”*, Cacucci, Bari, 495-506.

Zani S. (1993), “Classificazioni di unità territoriali e spaziali”, in S. Zani, *Metodi statistici per le analisi territoriali*, FrancoAngeli, Milano, 93-121.

Zimmerman J.H. (1976), “Description and optimization of fuzzy system”, in *J. Gen. Syst*. 2, 209-215.

Zimmerman J.H. (1978). Fuzzy programming and LP with several objective function, in J. Fuzzy Sets Syst. 1(1), 45-55.

Zimmermann (1976), "Description and optimization of fuzzy systems", Int. J. GEN System 2, 209-215.

Zimmerman (1985), *Fuzzy Sets Theory and its Application*, Kluver, Dordrecht.