

UNA RAPPRESENTAZIONE MULTIDIMENSIONALE DELLA QUALITÀ DELLA VITA
DELLE PROVINCE ITALIANE: CONFRONTO TRA RETI NEURALI SELF-
ORGANIZING MAP E MULTIDIMENSIONAL SCALING

Giovanni RABINO¹ e Nicola LONARDONI²

1 Di.A.P. - Dipartimento di Architettura e Pianificazione, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133, Milano

2 D.A.P.T. – Dipartimento di Architettura e Pianificazione Territoriale, Università degli Studi di Bologna, Facoltà di Ingegneria, V.le Risorgimento 2, 40135 Bologna

SOMMARIO*

Questo lavoro si propone di descrivere la multidimensionalità della qualità della vita attraverso una rappresentazione alternativa a quella della classificazione in graduatoria. In particolare si useranno i dati raccolti e divulgati da “Il Sole 24 ore”, relativamente alle Province Italiane, al fine di sperimentare l’applicazione di due tecniche in grado di cogliere non tanto una graduatoria assoluta, ma un posizionamento relativo delle osservazioni studiate. Si vogliono mettere a confronto due strumenti matematico-statistici: il Multidimensional Scaling e le Reti Neurali Self-Organizing Map.

Lavoro, sicurezza, salute, istruzione, ambiente e tenore di vita sono alcune delle tematiche che, secondo gli studi pubblicati, fanno di una regione semplicemente antropizzata un territorio di qualità. La rappresentazione attraverso una classifica è una scelta che si è rivelata assai utile sotto il profilo della comunicabilità dei risultati. Tuttavia, sul piano analitico e comparativo, il rischio è quello di considerare confrontabili unità completamente diverse tra loro. Una mappatura multidimensionale supera questo limite includendo gradi di libertà aggiuntivi che concedono maggiori possibilità di diversificazione.

* Sebbene questo studio sia il risultato di un lavoro congiunto, i paragrafi 1, 1.1, 2.1 e 2.2 sono da attribuirsi al lavoro di Giovanni Rabino; i paragrafi 1.2, 2, 2.3 e 2.4 sono dovuti al contributo di Nicola Lonardoni. E’ stato prezioso l’apporto di Stefania Lecchi e Angela Levanto nella redazione della parte tecnica.

1 TECNICHE DI RAPPRESENTAZIONE DELLA REALTÀ

Lo scopo di questo lavoro è quello di approfondire la conoscenza su alcuni strumenti analitici in grado di esplorare la realtà multidimensionale. Si vuole ricercare un metodo che tenga conto della realtà complessa, in grado nel contempo di superare i limiti di una classificazione per graduatoria, rigida e attraverso la quale non è possibile leggere una diversificazione.

Ciò che va maggiormente sottolineato è il fatto che nella varietà delle modalità con cui viene raggiunto un certo livello qualitativo, sono celate le relazioni tra i fattori che ne determinano successo o svantaggio. La comprensione di questi meccanismi e l'identificazione di questi fattori sono di fondamentale importanza per le scelte dei decisori.

Nello specifico si è scelto di applicare ai dati raccolti e pre-elaborati da “Il Sole 24 Ore”, due metodi di Feature Mapping. Il primo chiamato *MultiDimensional Scaling*, deriva dalla statistica tradizionale e rappresenta uno strumento basato sulla distanza relativa tra le osservazioni. E' questo elemento che va a garantire la permanenza della relazioni topologiche benché venga approssimata, a meno di un coefficiente chiamato *Stress*, la proiezione di uno spazio n -dimensionale su di un piano bi -dimensionale. Il secondo metodo deriva dalla teoria delle Reti Neurali ed è chiamato *Self-Organizing Map*. In questo caso la distanza euclidea è utilizzata per il confronto tra le osservazioni reali ed un insieme di punti n -dimensionali virtuali associati ai punti di una mappa. Il particolare algoritmo adattivo che auto-organizza il posizionamento dei punti prevede al suo interno l'applicazione di regole di prossimità in grado, anche in questo caso, di mantenere i legami topologici.

Fanno parte della famiglia delle tecniche di Feature Mapping le procedure che assegnano ad un oggetto, avente un numero indefinito n di proprietà, un punto di coordinate note in uno spazio di destinazione, o mappa. In generale, la dimensionalità dello spazio in cui giace la mappa è arbitraria; spesso viene utilizzato un piano. L'utilità di questo procedimento si manifesta soprattutto nel momento in cui ci si trova ad analizzare un collettivo appartenente ad uno spazio a dimensionalità molto elevata.

Infatti, nel caso di osservazioni composte da due a quattro variabili è in generale possibile definire univocamente una relazione per poter rappresentare le osservazioni in uno spazio che ne aumenti la leggibilità. In altre parole è possibile definire una relazione tra uno spazio vettoriale n -dimensionale ed uno $(n-1)$ -dimensionale. Tale relazione è identificata univocamente da una trasformazione detta “proiezione” ed in particolare, per la geometria analitica, lo spazio di destinazione è chiamato spazio proiettivo.

Tra spazi per i quali la differenza tra le dimensionalità è maggiore di 1, le trasformazioni che sono in grado di definire una corrispondenza sono invece infinite. Per questo motivo è necessario usufruire di strumenti in grado di predisporre una mappa leggibile, ottimizzando la conservazione delle relazioni topologiche tra gli oggetti. Il funzionamento di questo metodo è

di lettura immediata. Due osservazioni che vengono poste vicine o addirittura sovrapposte nello stesso punto, avranno caratteristiche molto somiglienti o addirittura coincidenti. Viceversa, due osservazioni poste in posizioni distanti avranno configurazioni molto dissimili. Questa operazione comporta necessariamente un certo grado di approssimazione e distorsione. I rapporti topologici tra le osservazioni sono conservati, ma ciò non implica la conservazione delle distanze reali. Naturalmente, valgono tutte le considerazioni effettuabili sui metodi di calcolo della somiglianza. Sebbene alla base di entrambi i metodi che presenteremo vi sia in la distanza euclidea, il calcolo di questa misura verrà applicato in due momenti operativi differenti. In un caso andrà a definire una corrispondenza più stretta tra gli aspetti quantitativi, sottolineando le componenti principali e evidenziando la corrispondenza tra casi quantitativamente più simili. Nell'altro caso, quello delle SOM, verrà preso maggiormente in considerazione il confronto tra i pattern, preservando le stesse relazioni ma con una modalità differente.

1.1 IL MULTI-DIMENSIONAL SCALING - MDS

Questo metodo è largamente usato in scienze come il marketing o nella psicometria, quale metodo di analisi dei dati. Come già accennato, l'obiettivo del Multidimensional Scaling è quello di fornire una rappresentazione grafica, in uno spazio di piccole dimensioni, di un set di p oggetti mantenendone i rapporti di prossimità (i.e. similarità, dissimilarità, distanze). Nel caso in cui avessimo una matrice in cui le distanze tra alcuni punti su una superficie siano date (ad esempio: le città di un Paese), il metodo MDS permetterebbe di ricostruire esattamente la mappa dei punti, eventualmente applicando delle trasformazioni come simmetrie rotazioni e/o traslazioni. Nel caso più generale, invece, questo metodo ci consente di costruire una mappa di una serie di p oggetti, partendo dalle distanze reciproche date in una matrice di dissimilarità/similarità, in uno spazio k -dimensionale, che in genere è 2D o 3D.

Operativamente, l'MDS è un algoritmo iterativo atto alla ricerca di un set di vettori in uno spazio k -dimensionale in modo tale che la matrice delle distanze euclidee tra i vettori risultanti corrisponda il più possibile alla matrice di dissimilarità iniziale, fornita come input al programma, a meno di un errore di approssimazione calcolato con una funzione chiamata stress.

Il grado di corrispondenza tra le distanze tra i punti suggerite dalla mappa del MDS e la matrice di dissimilarità immessa dall'utente è misurato da una funzione di stress: minore è lo stress, migliore è la rappresentazione. Tale funzione viene anche chiamata "Kruskal Stress", ed è data attraverso la formula seguente:

$$Stress = \sqrt{\frac{\sum \sum (f(x_{ij}) - d_{ij})^2}{\sum \sum d_{ij}^2}} \quad (1.1)$$

dove d_{ij} sono le distanze euclidee tra i punti i e j sulla mappa, e $f(x_{ij})$ è una funzione dei dati di input. Quando la mappa prodotta dal MDS riproduce perfettamente i dati di input, lo stress è nullo. Da un punto di vista matematico, valori di stress non nulli si hanno solo per insufficiente dimensionalità; perciò, per ogni set di dati, è impossibile rappresentare perfettamente i dati in 2 o altro numero ridotto di dimensioni. D'altra parte, un set di p osservazioni è perfettamente rappresentato usando $(n-1)$ dimensioni. Quindi, all'aumentare del numero di dimensioni, lo stress può diminuire o rimanere per lo più costante: per questo ci aspettiamo che esso sia una funzione decrescente con la dimensione di rappresentazione. Non è necessario che la mappa MDS abbia stress nullo per essere utile; una certa distorsione è sicuramente tollerabile.

1.2 LE SELF-ORGANIZING MAPS - SOM

Nei primi anni '80 Teuvo Kohonen diede un prezioso contributo alla realizzazione di questo tipo di Rete Neurale. Questa architettura viene utilizzata per molti scopi: il *Data Clustering*, la *Pattern Recognition* ed il *Feature Mapping*. Per ognuna di queste applicazioni l'algoritmo, l'apprendimento e gli elementi costitutivi non cambino, ciò che è differente è solamente l'approccio interpretativo dei risultati.

Il Feature Mapping effettuato da una SOM viene anche chiamato in letteratura Kohonen Feature Mapping. L'esercizio qui riportato è proprio un esempio di questa tecnica.

In primo luogo va considerato lo spazio vettoriale definito dalle p osservazioni studiate e dalle n proprietà di queste. Questo insieme costituisce il *Training Set* della Rete Neurale. In un secondo tempo va definito lo spazio in cui si vogliono rappresentare questi casi, ovvero la giacitura della mappa. Inoltre, poiché la mappa è rappresentata da un numero finito di punti biunivocamente corrisposti da un insieme di unità di elaborazione, deve essere definita la quantità di posizioni potenzialmente assumibili dalle osservazioni. Considerando la possibilità di un collettivo completamente eterogeneo, in numero minimo di posizioni va considerato pari a p ; ovvero al numero totale delle osservazioni stesse.

L'insieme di unità di elaborazione corrispondenti ai punti della mappa è detto Strato di Kohonen, il quale, per quanto detto sopra, possiede un numero $q > p$ elementi \underline{w} . Ogni elemento, chiamato anche Processing Element (PE), consta in un vettore n -dimensionale di pesi inizializzati in modo randomico.

Riassumendo avremo che ogni elemento PE si compone di un numero n di variabili w_i e di una coppia di coordinate (x_1, x_2) che ne identificano la posizione sulla mappa.

Ora è possibile avviare il processo di apprendimento non supervisionato. Questo è un processo iterativo che prevede la presentazione ciclica del Training Set alla SOM, la quale opera su un elemento alla volta un insieme di operazioni in più fasi. In un primo momento la

SOM verifica la dissomiglianza tra ogni PE, o meglio il vettore di pesi ad esso associato, e l'osservazione attiva. Una volta esauriti i PE, viene eletto quello vincitore, definito come *Winner Unit*, secondo questi criteri:

$$WinnerUnit = PE_s \Leftrightarrow C_s = \max_{r \in [1...m]} \left\{ C_r = X \cdot W_r = \sum_{i=1}^n x_i \cdot w_{ri} \right\} \quad (1.2)$$

$$WinnerUnit = PE_s \Leftrightarrow D_s = \min_{r \in [1...m]} \left\{ D_r = \|X - W_r\| = \sqrt{\sum_{i=1}^n (x_i - w_{ri})^2} \right\} \quad (1.3)$$

La Winner Unit è così definita come l'unità PE il cui vettore associato è maggiormente rappresentativo dell'osservazione corrente. La cella ad esso associata è a sua volta, e per il momento, la posizione in mappa dell'osservazione corrente.

Prima di passare all'osservazione successiva, i pesi di ogni PE vengono corretti secondo questa regola generale:

$$\Delta W(D) = h(D) \cdot (X - W_r) \quad (1.4)$$

dove $h(D)$ rappresenta una funzione di definizione del vicinato. Solitamente, quest'ultima è la funzione gaussiana:

$$h(D) = He^{-\frac{D^2}{2s^2}} \quad (1.5)$$

D rappresenta la distanza che intercorre tra i punti della mappa. Pertanto, più il PE che sta per essere corretto è prossimo al PE vincitore e maggiormente saranno avvicinati i propri pesi a quelli dell'osservazione corrente. Viceversa, più è lontano e più pesantemente saranno allontanati. In questo modo si garantisce che posizioni vicine sulla mappa rappresentino osservazioni simili nella realtà multidimensionale.

A questo proposito è necessario fare una precisazione: per la modalità dell'apprendimento, che distorce in modo indipendente la distribuzione delle variabili sulla mappa, non è possibile determinare una corrispondenza esatta tra le distanze calcolate sulla mappa e quelle reali. Le relazioni topologiche sono conservate non solo a meno di una approssimazione globale, come accade con il MDS, ma anche a meno di una distorsione locale della mappa. Questo fattore determina la differenza più sostanziale tra i due metodi.

Poiché la correzione dei pesi in funzione della distanza ha andamento gaussiano, è lecito affermare che la propagazione della correzione sull'intera mappa con queste modalità ha l'effetto di costruire una superficie per ogni i -esima variabile v_i nello spazio tridimensionale (x_1, x_2, v_i) che rispecchia l'andamento di v_i in relazione al mondo descritto dalle osservazioni

considerate. Questo passaggio è il carattere distintivo delle S.O.M., quello in cui emerge la sua particolarità. Questo metodo permette di ridisegnare, ciclo dopo ciclo, lo spazio di destinazione e la trasformazione che permette di rappresentare un insieme di p osservazioni n -dimensionali in 2 sole dimensioni. Il piano che verrà a formarsi sarà quindi costituito da posizioni contigue che compongono percorsi di transizione tra le varie modalità di cui è composta la realtà. Si vedrà in seguito come ciò produca informazioni leggibili non solo dal punto di vista della classificazione, ma anche dal punto di vista dell'analisi dei fattori caratterizzanti il fenomeno e nella costruzione di potenziali sentieri di trasformazione tra una configurazione ed un'altra.

2 UN ESEMPIO APPLICATIVO

Il tema della qualità della vita è sempre stato, e lo è tuttora, un argomento di fondamentale importanza in diverse fasi dei processi di decisione e di valutazione. Basti pensare che nell'identificazione della qualità sussiste implicitamente il riconoscimento dei fattori che ne determinano la presenza stessa e le modalità con cui essa è presente. In particolare, se si manifestano le condizioni di un assetto di qualità, ci si trova anche di fronte ad una configurazione per lo più risolta, per la quale gli elementi che ne definiscono la bontà dell'ambiente sono presenti ed in equilibrio. La qualità della vita, in passato confusa spesso con le possibilità materiali offerte dalla ricchezza economica, si compone invece di una molteplicità di aspetti che vanno dalla percezione della sicurezza all'accesso alla conoscenza, dalla salute al benessere economico, dalla casa ai mezzi di trasporto e comunicazione.

La vera difficoltà consiste nell'estrarre questi elementi dal rapporto peculiare che esiste tra un ambiente e la vita in esso contenuta. Un rapporto peculiare al punto da essere inteso quasi come univoco; complesso nella commistione di cause-effetti circolari che inducono un processo di auto-organizzazione nel sistema locale. In questo contesto ci si può domandare se esiste una chiave di lettura comune a questi oggetti unici, se esiste un metodo in grado di leggere una realtà complessa e di restituire informazioni utili alla definizione di strategie di sviluppo, di vantaggio competitivo, di intervento migliorativo.

Riassumendo in una semplificazione, possiamo dire che la qualità della vita in un luogo si compone della risposa che una situazione dà ai canoni di qualità locale e ai canoni di qualità globale. I primi permettono di soddisfare condizioni locali di necessità, i secondi sono legati a due fattori: le relazioni di scambio con altre realtà e le potenzialità attrattive di uomini ed investimenti. E' facile desumere che l'assenza delle prime pone un forte ostacolo alla persistenza delle seconde.

È proprio in funzione di queste considerazioni che ogni anno vengono redatte le classifiche delle Province Italiane. Autorevoli fonti divulgative, quali "Il Sole 24 Ore" considerano diversi aspetti della vita sociale ed economica delle Province e ne associano un punteggio.

Una buona Provincia è sicura, garantisce un lavoro, tutela la salubrità dell'ambiente e la serenità della società che vi risiede. Una buona Provincia è un luogo in cui vivere risulta sicuro, economico, piacevole e gratificante; una buona Provincia è un luogo in cui è più facile che un investimento vada a buon fine. Queste graduatorie mettono a confronto le diverse Province stilando le classifiche su scale relative e aggregando i punteggi ottenuti in relazione ai differenti aspetti considerati. La classificazione così prodotta ha un carattere monodimensionale; non permette cioè di constatare la diversificazione delle osservazioni, una diversificazione imposta non solo dalla differente configurazione qualitativa, ma implicitamente anche dalle differenti condizioni ambientali.

2.1 I DATI DI PARTENZA

I dati di partenza della sperimentazione effettuata sono quelli divulgati da “Il Sole 24 Ore” e disponibili al libero download attraverso il sito del quotidiano. Riguardano i temi degli Affari e Lavoro; la Criminalità; la Popolazione; i Servizi e l'Ambiente; il Tempo Libero ed il Tenore di Vita. Ognuno di questi temi viene descritto da 6 proprietà e dai rispettivi valori. Pertanto ogni Provincia risulta associata ad un vettore di 36 componenti. Per contestualizzare la sperimentazione in un progetto di approfondimento più ampio, che tenesse conto del fattore temporale, si è considerata una base di dati avente due annate di riferimento: il 2002 ed il 2003. Per rendere omogenei i dati in oggetto si è proceduto ad intersecare l'insieme delle variabili presenti nelle due annate per ottenere un sottoinsieme comune di variabili che permettesse, quindi, il confronto diretto tra 2002 e 2003 appunto. Questo sottoinsieme è costituito da 25 elementi. In definitiva, le Province si sono considerate associate a vettori di 25 componenti. La riduzione non è stata omogenea per ogni area tematica, pertanto, ai fini della lettura è necessario tenere conto del fatto che potrebbero esserci delle distorsioni dovute alle aree tematiche con un numero maggiore di variabili considerate. Queste note sono doverose dal punto di vista metodologico generale, ma non incidono sullo scopo generale di questo esercizio.

Inoltre, i limiti del software dimostrativo impiegato nella realizzazione del MDS, hanno imposto la riduzione del numero di province a 35, estratte ordinatamente dal totale.

Il metodo comporta necessariamente una standardizzazione del dato. Questa è stata effettuata tramite una trasformazione con traslazione per mezzo della formula:

$$\overline{X}_i = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \quad (2.1)$$

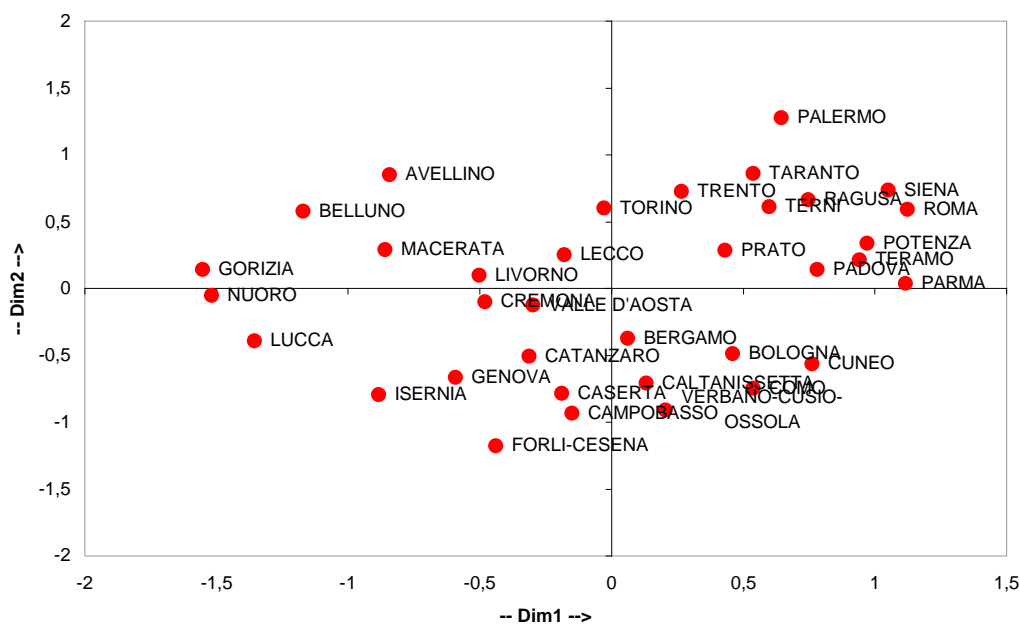
Come si è più volte ripetuto all'interno di queste pagine, lo scopo principale è quello di dare una rappresentazione bidimensionale della realtà descritta da queste 35 Province. In effetti, entrambe i metodi hanno lo scopo di ricostruire in un sistema di coordinate piane, l'insieme di

posizioni relative che queste osservazioni assumono nello spazio a n dimensioni. In altre parole tentano di riportare su un piano le relazioni topologiche tra i casi in esame.

L'apporto informativo che ci si propone di dare attraverso un'analisi di questo tipo, riguarda principalmente due obiettivi. Il primo è quello di riconoscere delle tipologie di Province. Se dovessero risultare delle concentrazioni elevate di gruppi di osservazioni, l'analisi alla quale si sarebbe portati evidenzierebbe la somiglianza interna, l'omogeneità dell'insieme individuato. Al contrario una distribuzione per lo più omogenea sul piano porterebbe a concludere che le osservazioni hanno configurazione essenzialmente eterogenee. Il secondo obiettivo, sicuramente più complesso, sarebbe quello di cogliere le modalità di diversificazione dei casi. L'individuazione di fattori distintivi o migliorativi è in fondo lo scopo applicativo di questi strumenti all'analisi a supporto delle decisioni. Nell'avanzamento della ricerca, si profila il raggiungimento di un terzo obiettivo generale che riguarda l'applicazione di queste analisi a sistemi di questo tipo che si pongono in divenire.

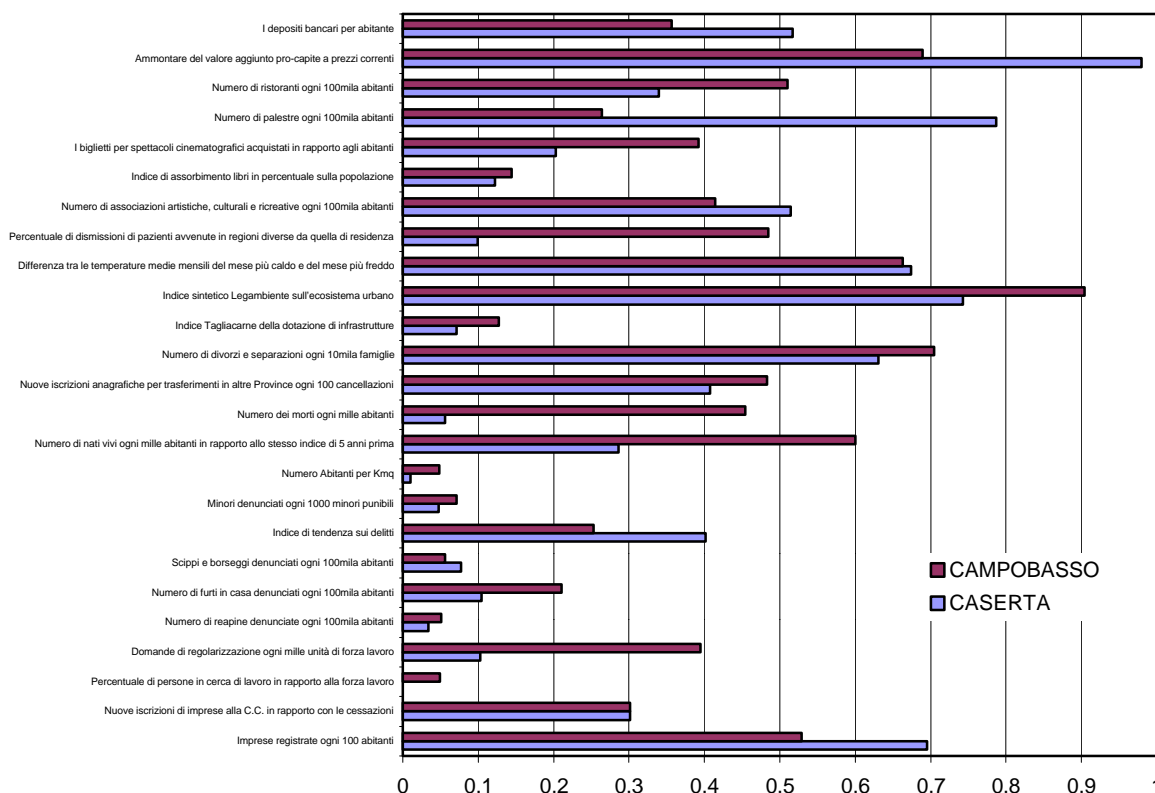
2.2 I RISULTATI OTTENUTI DA MDS

Figura 1 - Applicazione dell'MDS alle 35 Province selezionate (Stress: 0,219)



L'applicazione dell'MDS ha individuato una distribuzione delle osservazioni in una nuvola di punti piuttosto compatta, omogenea il cui baricentro è individuato da Aosta. Pochi casi tendono ad avvicinarsi significativamente, i più evidenti sono Caserta e Campobasso, Potenza e Teramo, Siena e Roma. Confrontando i vettori originali, ad esempio della prima coppia citata, è possibile notare una forte tendenza dei pattern alla sovrapposizione; una tendenza coerente con quanto rilevato dallo strumento.

Grafico 1 - Cfr. Campobasso vs. Caserta

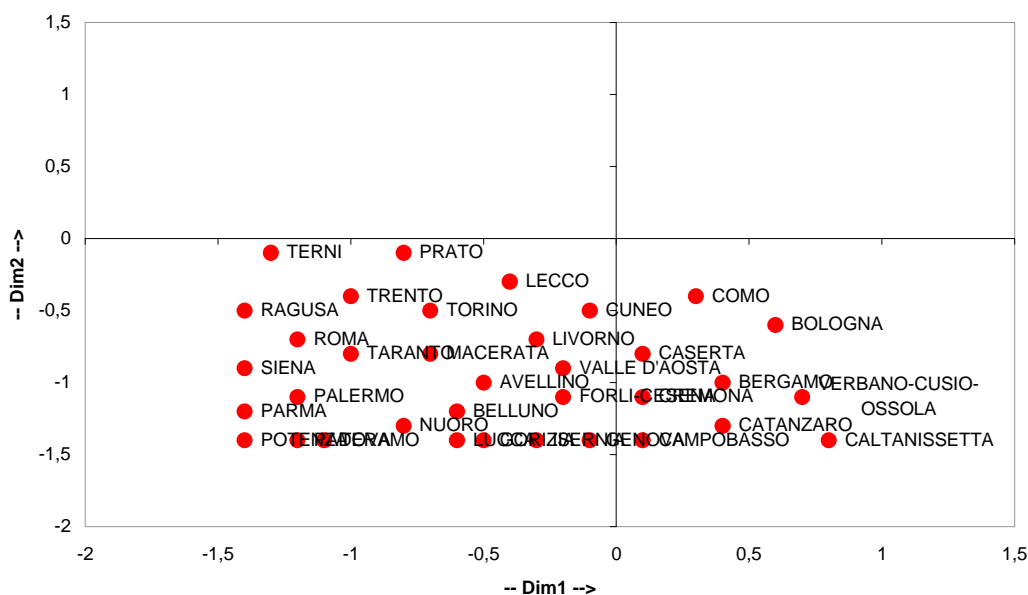


2.3 I RISULTATI OTTENUTI CON S.O.M.

L'applicazione dell'algoritmo S.O.M. ha prodotto un risultato differente sotto molti aspetti. Per rendere confrontabili i risultati ultimi, si è proceduto ad una scalatura della mappa ottenuta in output e una traslazione degli assi. Va inoltre considerato che il piano di proiezione utilizzato dalla S.O.M., è un piano discreto. Si compone infatti di nodi equidistanti i quali vengono associati alle osservazioni al termine del processo di apprendimento.

Come si evince dalla figura, la S.O.M. costruisce una nuvola più compatta di quella definita dall'MDS. Anche in questo caso Aosta assume una posizione centrale rispetto alle 35 Province considerate. La distribuzione generale è diversa da quella della precedente applicazione, benché nel dettaglio si riscontrino moltissime corrispondenze. Si noti che nella nuvola, globalmente molto compatta, non esistono tuttavia casi di sovrapposizione, cosa che equivarrebbe a definire una generale somiglianza tra gli oggetti della stessa natura, ma una certa eterogeneità interna che ne sottolinea chiaramente una forte differenziazione. In altre parole, lo strumento ha identificato e amplificato le caratteristiche strutturali degli oggetti studiati concentrandoli in una porzione limitata dello spazio totale su cui poteva distribuirli, ma ha contemporaneamente riscontrato una molteplicità di configurazioni differenti tra loro in modo peculiare.

Figura 2 - Applicazione dell'algoritmo SOM ai dati delle 35 Province selezionate



Il caso riportato precedentemente, di Campobasso e Caserta, non è stato riconosciuto dalla S.O.M.. Tra le due Province, infatti, se ne interpone una terza: Cremona.

Tabella 1 – Matrice delle distanze ridotta alle Province di Campobasso, Caserta e Cremona.

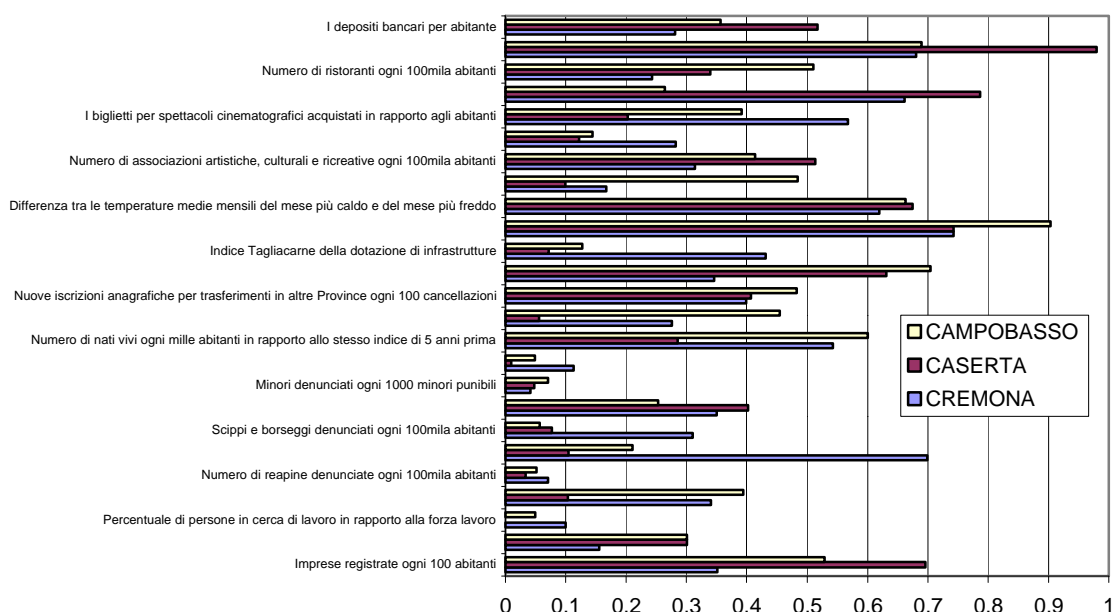
	CAMPOBASSO	CASERTA	CREMONA
CAMPOBASSO	0,0000	4,1328	4,2240
CASERTA	4,1328	0,0000	4,0969
CREMONA	4,2240	4,0969	0,0000

Si nota dalla *Tabella 1* che le tre osservazioni risultano essere approssimativamente equidistanti, ma Caserta e Cremona risultano essere le più vicine. Questo fatto è coerente con i risultati ottenuti dalla SOM, ma non con quelli ottenuti dall'MDS, che, invece predilige rappresentare la vicinanza tra Caserta e Campobasso. Da quanto emerge da questa matrice, ci si aspetterebbe tuttavia di trovare questi elementi rappresentati come vertici di un triangolo pressoché equilatero. Va tenuto conto tuttavia che entrambe le metodologie non elaborano solamente queste tre osservazioni, bensì vanno a forza una distribuzione di 25 punti. Questi scarti, dovuti alle modalità di elaborazione dei due algoritmi, manifestano i limiti dell'approssimazione necessariamente presente in questo tipo di strumenti.

Dall'altro lato, se ora confrontiamo i pattern delle tre osservazioni nel *Grafico 2* si evince che Cremona manifesta un pattern che talvolta si presenta simile a quello di Campobasso e talvolta a quello di Caserta, ma sono pochi i punti in cui si discosta da entrambe le

osservazioni.

Grafico 2 - Campobasso, Caserta e Cremona



Questo elemento sottolinea la caratteristica della SOM di cogliere le similitudini tra i pattern prediligendo la configurazione d'insieme alla sovrapposizione parziale delle proprietà.

2.4 L'INCORNTRO DEI DUE METODI

Una ulteriore sperimentazione è stata effettuata applicando l'MDS ai vettori dei pesi relativi alle Winner Units elette dalla SOM. In altre parole si è applicata la scalkatura dimensionale alla rappresentazione interna che la SOM si è creata durante l'apprendimento.

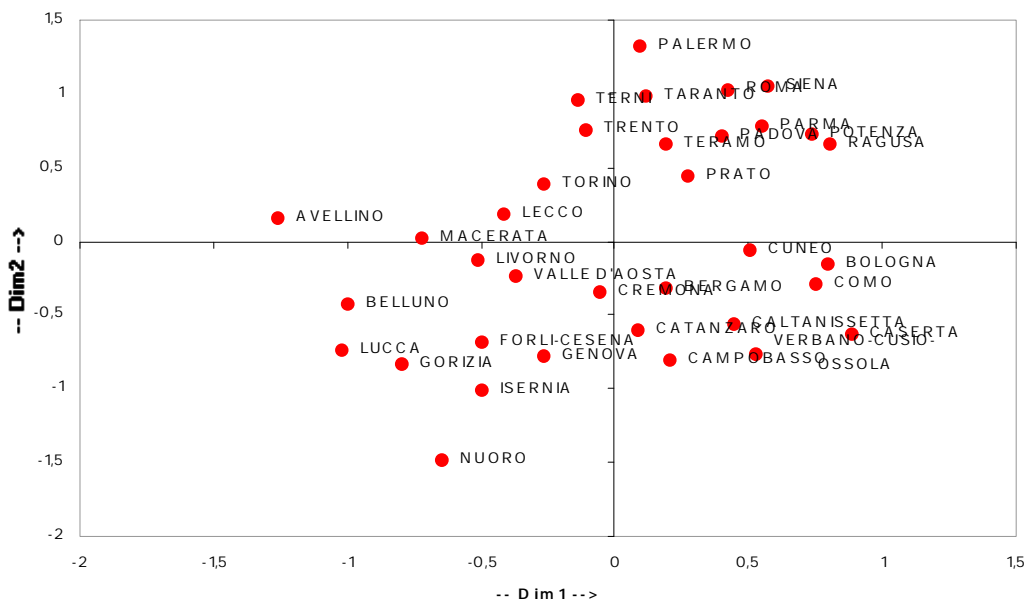
La nuvola di punti che viene qui descritta ha la stessa conformazione degli altri casi. In questa mappa, Aosta risulta leggermente decentrata rispetto al centro della nuvola. Inoltre, lo Stress è molto minore rispetto alla prima sperimentazione. Ciò è probabilmente dovuto al fatto che parte dell'approssimazione è già stata assorbita dal pre-processamento effettuato tramite la SOM.

In questo caso è possibile riconoscere le tre Province che abbiamo preso in esame (Campobasso, Caserta e Cremona) disposte come vertici di un triangolo.

Confrontando la prima e l'ultima mappa emerge evidente una forte coerenza tra le due rappresentazioni topologiche, ma non una corrispondenza locale nella distribuzione dei punti. La Provincia di Bologna ad esempio risulta essere molto più "vicina" alla Provincia di Bergamo di quanto non lo fosse nella prima rappresentazione. A questo proposito va sottolineato che sia l'approssimazione data dal MDS che quella data da processamento della SOM partono dall'inizializzazione casuale di variabili che vengono iterativamente corrette o

adattate. Questo punto di partenza nell'elaborazione produce una differenza più o meno trascurabile nei risultati.

Figura 3 - Algoritmi MDS applicato ai vettori dei pesi che il procedimento della SOM ha associato alle 35 Province selezionate (Stress: 0,191)



In questa rappresentazione è necessario mettere in evidenza due elementi. In primo luogo esiste la tensione a marcare delle aggregazioni. Nel primo quadrante, in alto a destra, sono presenti le stesse osservazioni riportate nel medesimo quadrante dall'applicazione diretta dell'MDS. In quest'ultimo test tuttavia, la densità dei punti aumenta, come aumenta quella nel quarto quadrante.

Un secondo punto è che nell'elaborazione della SOM, i risultati vengono comunque espressi in una suddivisione discreta dello spazio. Applicando l'MDS ai vettori corrispondenti alla Winner Unit è possibile avere una mappatura nel continuo di ciò che costituisce la rappresentazione interna che l'algoritmo neurale elabora in fase di apprendimento.

3 CONCLUSIONI

Dal confronto di queste tecniche sono emersi diversi elementi. Innanzitutto di carattere metodologico-operativo, laddove è necessario ricostruire ed analizzare il costrutto topologico di un insieme di elementi.

Infatti, l'MDS pone dei vincoli geometrici che determinano una ricostruzione molto rigida delle relazioni tra gli oggetti. Questa rigidità ha come conseguenza un grande perdita di informazione.

L'algoritmo delle SOM non impone per contro delle restrizioni geometriche in questo senso. Tuttavia, il mantenimento di una maggiore apertura produce una perdita di rigidità che si paga nella leggibilità dei risultati. La tendenza dello strumento è infatti quella di rafforzare i legami

già forti e di distorcere lo spazio di destinazione in modo da evidenziare come maggiormente distanti legami già deboli. E' lecito affermare che in questo caso viene preservata la struttura topologica delle osservazioni, e questo è confermato dai risultati dell'applicazione dell'MDS ai pesi delle Winner Units.

In conclusione, le potenzialità di entrambi gli strumenti emergono in contesti diversi.

L'MDS si adatta con precisione ad applicazioni con un numero limitato di variabili e laddove si renda maggiormente necessaria una lettura più dettagliata di insiemi piccoli di osservazioni. Le SOM, data la loro versatilità, vanno considerate per analisi di più ampio respiro, dove la generalizzazione e l'approssimazione risultano più vantaggiose della rigidità. Con un collettivo molto ampio ed uno spazio a dimensionalità molto elevata, il grado di approssimazione che sarebbe ottenuto dal MDS sarebbe assai più elevato rispetto a quello ottenuto dalla SOM.

Vanno comunque sempre considerati i limiti interpretativi individuabili dagli strumenti. L'approssimazione della tecnica MDS, infatti, può produrre delle distorsioni locali fuorvianti. La generalizzazione della SOM pertanto produce una distorsione globale.

Sono molte le questioni e le problematiche ancora aperte. Questi esercizi rappresentano un passo avanti dell'approfondimento che questa ricerca propone. Le differenti tecniche hanno peculiarità importanti che vanno studiate e sfruttate al meglio. Inoltre, un ulteriore avanzamento si otterrebbe dal proseguire queste elaborazioni in funzione del fattore temporale ed evolutivo.

ABSTRACT

Inspired by the life's quality studies published annually from "Il Sole 24 Ore", this job wishes to describe the life's multidimensionality with an alternative representation. Above all, the selected data will be used to apply the two Feature Mapping based techniques to the Italian Provinces in order to give a relative positioning of the studied observations. Contemporarily, another goal of this paper is to compare the two mathematical-statistic tools: The Multidimensional Scaling and the Neural Networks Self-Organizing Map.

Work, safety, health, education, environment are some of the properties that make a quality region, according to the studies conducted by the "Il Sole 24 Ore". The classification is a choice that has revealed greatly useful under the communicability profile. It is founded on a weighed score system that allows the association of an observation to a position. On the base of the total score gotten, it is usually possible to measure the distance between two observations. Nevertheless, the use of this method implies the loss of many information about the matters of the this position. Therefore, this method doesn't individualize a measure of dissimilarity, so near observations have not necessarily comparable configurations. On the analytical-comparative level, the risk is to set in competition completely different units. For example, it is possible that regions with really different characteristics can obtain near positions. For instance, in the section "Business and Job", the provinces Ancona and Piacenza assume the same position in classification. Comparing the respective detailed configurations however, they result to be deeply different. The multidimensional mapping overcomes this limit and allows the observations "to deflect" in two or more different directions, including in this way the possibility to diversify their own similarity. This fact can be exasperated in the comparison between three equidistant provinces. In this case, for the geometry of the linear representation, it is impossible to have a complete information. Contrarily it would happen in a bi-dimensional space where this important information would easily be deducible.

Bibliografia

- Bolasco S., (1999); *Analisi multidimensionale dei dati. metodi, strategie e criteri*, Carocci, Roma.
- Buscema M. & Semeion Group, (eds.) (1999a); *Reti neurali artificiali e sistemi sociali complessi. Teoria – Modelli - Applicazioni; Volume I: - Teoria e Modelli*; Franco Angeli, Milano.
- Buscema M. & Semeion Group, (eds.) (1999b); *Reti neurali artificiali e sistemi sociali complessi. Teoria – Modelli - Applicazioni; Volume II: - Applicazioni*; Franco Angeli, Milano.
- Kohonen, T. (1995), *Self-Organizing Maps*. Springer, Berlin.
- Kruskal Joe B., (1964), Non metric multidimensional scaling: a numerical method. *Psychometria*, USA, vol.29. – p. 115-129.
- Occelli S., Novelli E. (1999), Profili descrittivi di distribuzioni spaziali: alcune misure di diversificazione. Atti della *XIX Conferenza Nazionale dell' Associazione di Italiana di Scienze Regionali (AISRE)*. A.I.S.Re., Milano. L'articolo è presente anche in: *Cybergeo*, 108, 1999.