Bollettino Unione Matematica Italiana

Sezione A – La Matematica nella Società e nella Cultura

Cristina Coppola

Misure di distanza e similarità in spazi di informazione

Bollettino dell'Unione Matematica Italiana, Serie 8, Vol. **10-A**—*La Matematica nella Società e nella Cultura* (2007), n.2, p. 207–210.

Unione Matematica Italiana

<http://www.bdim.eu/item?id=BUMI_2007_8_10A_2_207_0>

Bollettino dell'Unione Matematica Italiana, Unione Matematica Italiana, 2007.

Bollettino U. M. I.
La Matematica nella Società e nella Cultura
Serie VIII, Vol. X-A, Agosto 2007, 207-210

Misure di distanza e similarità in spazi di informazione

CRISTINA COPPOLA

1. – Introduzione.

Il lavoro di ricerca svolto nella tesi ha preso spunto da una interessante questione di "fondamenti della geometria": come si possa realizzare una teoria geometrica senza assumere la nozione di "punto" come primitiva. Tale questione è stata per la prima volta estensivamente esaminata dal filosofo e matematico A. N. Whitehead, il quale assume come nozione primitiva quella di "regione" ([6]) e definisce, invece, i punti tramite opportune successioni di regioni. Nella tesi ci si riferisce ad approcci di tipo metrico alla "geometria senza punti", partendo da una serie di articoli di G. Gerla (si veda ad esempio [4]). Inoltre, si interpreta la nozione di regione in termini di informazione "incompleta". Il diametro di una regione, infatti, viene interpretato come una misura dell'incompletezza dell'informazione: al diminuire del diametro, aumenta l'informazione, fino ad arrivare al "punto" (diametro nullo), che rappresenta informazione completa. Riguardo al trattamento di informazione, in letteratura la nozione di spazio metrico gioca un ruolo importante. In generale, gli oggetti da esaminare possono essere rappresentati come punti in uno spazio e la distanza tra punti rappresenta una misura della "dissimilarità" tra oggetti. Nella tesi si affronta la questione riguardante l'adeguatezza di tale nozione, in un contesto in cui non è disponibile un'informazione completa riguardo agli oggetti in esame. Per affrontare tale questione, si sono definite distanze verificanti assiomi più generali di quelli delle metriche in spazi di "regioni", anziché di punti.

Strettamente legate all'approccio metrico, si sono sviluppate anche nozioni derivanti dalle logiche a più valori, con particolare attenzione agli *ordinamenti fuzzy* ed alle *similarity*. Ricordiamo alcune nozioni base.

DEFINIZIONE 1. – *Data una t-norma $\otimes$ ed una relazione binaria fuzzy su un insieme S,* $ord : S \times S \to [0,1]$, *consideriamo le seguenti proprietà:*

(1) $ord(x,x) = 1$ *(riflessività)*
(2) $ord(x,y) = ord(y,x)$ *(simmetria)*
(3) $ord(x,y) \otimes ord(y,z) \leq ord(x,z)$ *($\otimes$-transitività)*
(4) $ord(x,y) = ord(y,x) = 1 \Rightarrow x = y$ *(antisimmetria)*

Diciamo che ord è

- *fuzzy preordine se soddisfa* (1) *e* (3),
- *fuzzy ordine se soddisfa* (1), (3) *e* (4),

- *fuzzy similarity se soddisfa* (1), (2) *e* (3),
- *fuzzy similarity stretta se soddisfa* (1), (2), (3) *e* (4).

Le nozioni di "distanza" e "similarità", sono chiaramente duali: nel confrontare oggetti in base alle proprietà che verificano, si può utilizzare sia una misura che riveli quanto sono "simili", sia una misura che ne evidenzi le differenze, scegliendo, a seconda delle situazioni, la più opportuna. Ad esempio, le nozioni di similarity ed ordini fuzzy, in ambito di logiche a più valori, sembrano più adatte in situazioni in cui gli oggetti in esame verificano proprietà "vaghe". Tale dualità, che viene formalizzata in modo preciso, è spesso utilizzata nella tesi e stabilisce un ponte tra nozioni di carattere metrico e nozioni di carattere logico.

2. – Risultati Principali.

Facendo riferimento alle dualità tra relazioni fuzzy e distanze, si è stabilito un collegamento tra spazi metrici "senza punti" e categorie di fuzzy set, coinvolgendo la nozione di similarity debole. Più precisamente, particolare attenzione è stata rivolta alle ultrametriche, spesso utilizzate in letteratura in processi di classificazione. Sono stati definiti, infatti, gli *spazi ultrametrici pointless*, che sono strutture $(R, \leq, \delta, ||)$, dove $(R, \leq)$ è un insieme ordinato di regioni, δ e $||$ sono due funzioni rispettivamente *order-reversing* ed *order-preserving* sulle regioni, verificanti opportuni assiomi ([1]). Dualmente, l'attenzione è stata rivolta a similarity relative alla t-norma del mi-nimo. Sono state definite relazioni tra le strutture in ambito metrico e quelle con fuzzy relazioni. In particolare, si è trovato un collegamento tra gli spazi ultrametrici pointless e strutture con una particolare relazione fuzzy, una semi-similarity con la t-norma di Gödel, chiamata *G-semisimilarity*. Questa relazione è stata esaminata anche da un punto di vista categoriale ([1]): una volta organizzata la classe degli spazi ultrametrici pointless in una categoria, sono stati definiti due funtori per metterla in relazione con la categoria di fuzzy set proposta da Höhle ([5]).

Il numero $\delta(x, y)$ è, in un certo senso, una distanza minima tra le regioni x ed y. Nella tesi si è coinvolta anche una distanza massima, giungendo alla nozione di *distanza approssimata*. Le distanze approssimate estendono la nozione usuale di distanza, tenendo conto dell'errore che si può commettere a causa della non completezza dell'informazione che si ha a disposizione. Per distanza approssimata si intende una funzione che associa ad ogni coppia di oggetti un intervallo del campo dei numeri reali, intervallo che fornisce il range in cui può variare il valore effettivo della distanza. Si perviene allora alla seguente teoria dove viene coinvolta una funzione "interval-valued" Δ e dove le operazioni e le relazioni coinvolte negli assiomi sono quelle definite nell'*interval analysis*.

DEFINIZIONE 2. – *Un interval semimetric space è una struttura* $(Re, \leq_{Re}, \Delta)$, *con* $(Re, \leq_{Re})$ *insieme parzialmente ordinato e* Δ *funzione definita in* $Re \times Re$ *ed a* *valori nell'insieme I degli intervalli chiusi sui reali, tale che, per ogni $x, y, z \in Re$, valgono i seguenti assiomi:*

A1) $\Delta(x,x) \cdot [0,1] = \Delta(x,x)$,
A2) $\Delta(x,y) = \Delta(y,x)$,
A3) $\Delta(x,y) - \Delta(z,z) \leq_I \Delta(x,z) + \Delta(z,y)$,
A4) $\Delta(x,y) - \Delta(x,y) \leq_I \Delta(x,x) + \Delta(y,y)$,
A5) $x \leq_{Re} x', y \leq_{Re} y' \Rightarrow \Delta(x,y) \subseteq \Delta(x',y')$.

Anche in questo caso, gli elementi $x \in Re$ sono interpretabili sia come regioni sia come "pezzi di informazione" non necessariamente completa. L'intervallo $\Delta(x,y)$ esprime una misura approssimata di quanto due pezzi di informazione x ed y siano vicini. Definiamo anche una *funzione peso* $p : Re \to \boldsymbol{R}^+$, come $p(x) = \pi_2(\Delta(x,x))$, dove π_2 è la seconda proiezione dell'intervallo $\Delta(x,x)$. Interpretiamo $p(x)$ come una misura della completezza di x. Modelli canonici della risultante teoria sono ottenuti, a partire da classi di sottoinsiemi limitati di spazi pseudometrici, definendo distanze-intervallo tramite la *minima* e la *massima* distanza tra sottoinsiemi. Tuttavia, interessante è anche la considerazione di modelli non canonici. Ad esempio, sono definite distanze approssimate tra fuzzy set in due modi. Nel primo, si utilizza la nozione di *tagli* di un fuzzy set; nel secondo modo, ci si riferisce alla nozione di *ipografo* di un fuzzy set. Si ottengono ancora modelli non canonici definendo distanze approssimate tra *rough set*. Infine, viene proposta un'applicazione di tali distanze approssimate ad una procedura di clustering agglomerativo gerarchico, mediante un algoritmo che utilizza, appunto, una distanza approssimata tra cluster ([3]).

Tra le distanze verificanti assiomi più deboli di quelli delle metriche, l'attenzione è stata focalizzata anche su distanze non simmetriche, le quasi-metriche, e principalmente sulla nozione duale, quella di fuzzy inclusione. Il modello canonico che ispira tale approccio è sempre legato alla geometria senza punti, in quanto un tipico esempio di quasi-metrica è fornita dalla nozione di eccesso di una regione rispetto ad un altra. Le nozioni che si elaborano permettono di sviluppare una teoria dei punti fissi che unifica ed estende sia la teoria dei punti fissi in insiemi ordinati che quella in ambito metrico. Più precisamente, dopo aver definito le nozioni di completezza, di mappe contrattive e continue in strutture con ordini fuzzy, si provano teoremi di punto fisso ([2]). Un teorema di esistenza è il seguente:

TEOREMA 1. *Data una struttura S con un fuzzy ordine ord completa ed una funzione continua e contrattiva f della struttura in sè, se esiste un elemento x_0 in S tale che $ord(x_0, f(x_0)) \neq 0$, allora la successione $f^n(x_0)$ converge ad un limite l, che è un punto fisso per f.*

Si esamina la possibilità di applicare le nozioni proposte nell'ambito della programmazione logica, per la ricerca di punti fissi per l'operatore di conseguenza immediata. Ciò è particolarmente importante per i programmi con negazione, in cui l'operatore di conseguanza immediata può non essere monotono, cosa che rende non utili gli usuali teoremi di punto fisso in insiemi ordinati. Inoltre, teoremi di punto fisso in strutture con ordinamenti fuzzy possono risultare utili nel caso di programmazione logica fuzzy, anche in presenza di monotonia. Infatti, il processo che conduce al fuzzy set di conseguenze, a partire da un fuzzy set di ipotesi, avviene in un ambiente "continuo". Ne segue che non sempre risulta possibile ottenere un output esatto dopo un numero finito di passi e che, quindi, è necessario accettare risultati approssimati. Ciò è consentito nell'ambiente degli ordinamenti fuzzy, che permette non solo di calcolare punti fissi "approssimati", ma anche di controllare l'accuratezza dell'approssimazione.

RIFERIMENTI BIBLIOGRAFICI

[1] COPPOLA C., GERLA G., PACELLI T., *Point-free Ultrametric Spaces and the Category of Fuzzy Subsets,* Proceedings of the Tenth International Conference IPMU 2004, **1** (2004), 503-510.
[2] COPPOLA C., GERLA G., PACELLI T., *Convergence and fixed points by fuzzy orders,* iniviato a rivista.
[3] COPPOLA C., PACELLI T., *Approximate distances, pointless geometry and incomplete information,* Fuzzy Sets and Systems, **157** (2006), 2371-2383.
[4] GERLA G., *Pointless Metric Spaces,* Journal of Symbolic Logic, **55**, **1** (1990), 207-219,
[5] HÖHLE U., *Presheaves over GL-monoid,* In Nonclassical Logics and Their Applications to Fuzzy Subsets (U. Höhle, E. P. Klement, eds), Kluwer (1995), 127-157.
[6] Whitehead A. N., Process and Reality, McMillian, N.Y., (1929)

Dipartimento di Matematica e Informatica, Università degli Studi di Salerno
e-mail: ccoppola@unisa.it
Dottorato in Scienze Computazionali e Informatiche
(sede amministrativa: Università degli Studi di Napoli Federico II) - Ciclo XVIII
Direttore di ricerca: Prof. Giangiacomo Gerla, Università degli Studi di Salerno