

Il teorema di Takens: varianti ed applicazioni

Antonio De Capua

Scuola Normale Superiore

Dinamica e Serie Temporalì, 8 giugno 2011

Introduzione

- Il **teorema di Takens** (1980) “autorizza” la ricostruzione della dinamica di sistemi di varia natura a partire da serie temporali.
- Nella sua forma standard esso è applicabile solo a sistemi **autonomi**, e all’atto pratico necessita di essere applicato in un modo “ottimizzato”.
- Vedremo quali sono i progressi che sono stati fatti in tempi recenti nel dimostrarne varianti o migliorarne le applicazioni.
- Parleremo dei risultati che dà sull’analisi dei dati provenienti da **fMRI**.

La situazione di interesse

Il teorema di Takens si applica a sistemi dinamici discreti.

La situazione di interesse

Il teorema di Takens si applica a sistemi dinamici discreti.

- lo spazio delle fasi è una varietà differenziabile compatta M
- l'evoluzione del sistema è data da una mappa $f \in \text{Diff}(M)$
- f in generale cambia ad ogni passo in base ad uno o più parametri. Se ciò non accade il sistema è **autonomo**.

Si vogliono ricostruire le proprietà del sistema dinamico in base ai valori assunti nel tempo da un **osservabile**, cioè un'applicazione $\phi : M \rightarrow \mathbb{R}$.

Enunciato del teorema

Per ora ci occupiamo di sistemi autonomi.

La **delay map** associata a f e ϕ , fissato un $d \geq 0$, è

$$\begin{aligned}\Phi_{f,\phi} : M &\rightarrow \mathbb{R}^d \\ x &\mapsto (\phi(x), \phi(fx), \dots, \phi(f^{(d-1)}x))\end{aligned}$$

Teorema (Takens)

Sia M^m una varietà compatta, e siano $r \geq 1$, $d \geq 2m + 1$. Allora esiste un sottoinsieme aperto denso di

$$(f, \phi) \in \text{Diff}^r(M) \times \mathcal{C}^r(M, \mathbb{R})$$

tali che la delay map $\Phi_{f,\phi}$ è un embedding.

Come si usa il teorema di Takens

Supponiamo di avere una serie temporale data da

$$\phi_i = \phi(f^{(i)}x_0).$$

- Sia $F = \Phi_{f,\phi} \circ f \circ \Phi_{f,\phi}^{-1}$ la mappa coniugata a f su $\Phi_{f,\phi}(M)$.
- Detto $z_i = (\phi_i, \phi_{i+1}, \dots, \phi_{i+d-1})$ si ha allora $F(z_i) = z_{i+1}$.
- Avendo a disposizione una serie temporale abbastanza lunga, è possibile tentare un'approssimazione di F : quindi di descrivere proprietà del sistema dinamico in esame a meno di coniugio.

Problemi di applicazione

- L'ipotesi di sistema dinamico autonomo è molto restrittiva nella pratica.
- Nei fatti si hanno a disposizione serie temporali non infinite, e talvolta anche piuttosto corte.
- Senza conoscere il sistema dinamico a monte, non si sa neanche quanto deve essere alta la **dimensione di embedding** d .

Fissato un intero positivo τ (il **lag**), è possibile definire la delay map anche come

$$x \mapsto \left(\phi(x), \phi(f^{(\tau)}x), \dots, \phi(f^{((d-1)\tau)}x) \right)$$

ottenendo gli stessi risultati.

È indispensabile capire qual è il lag che dà i migliori risultati nell'embedding.

Mutua informazione

Siano S e Q due sistemi che presentano un ventaglio di output possibili $\{s_j\}$ e $\{q_j\}$ con probabilità note.

- L'entropia è la quantità $H(S) = \sum_j p_S(s_j) \log p_S(s_j)$
- La mutua informazione fra S e Q è

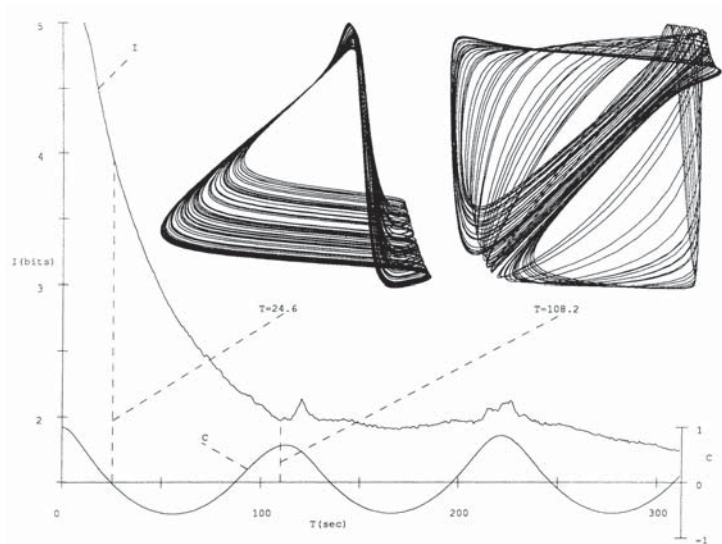
$$I(S, Q) = H(S) + H(Q) - H(S, Q)$$

La mutua informazione è una misura di quanto la conoscenza di uno dei due sistemi aiuta a “spiegare” l'altro.

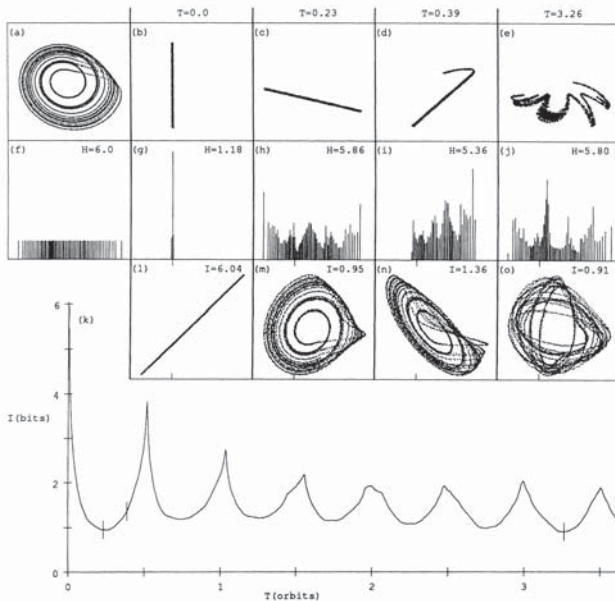
Rispetto alla più classica correlazione, ha il pregio di considerare anche i legami non lineari.

La scelta del lag

- Consideriamo come ventagli di output la serie $\{\phi_i\}$ e la serie ritardata $\{\phi_{i-\tau}\}$, e ne stimiamo quali sono le probabilità associate, sia disgiunte che congiunte.
- Sia I_τ la relativa mutua informazione.
- Una buona scelta per il lag τ è allora il primo punto di minimo locale di I_τ .



Embedding dell'attrattore di Roux.



Attrattore di Rössler

$$\begin{cases} \dot{x} = -z - y \\ \dot{y} = x + ay \\ \dot{z} = b + z(x - c) \end{cases}$$

I prodotti obliqui

- Modellizziamo un **sistema forzato** come un sistema M in cui la mappa f_y è determinata dallo stato y in cui si trova un altro sistema (autonomo) N , la cui evoluzione è descritta da $g \in \text{Diff}(N)$.
- Matematicamente, si potrebbe pensare $N \subseteq \text{Diff}(M)$.
- L'evoluzione congiunta nel tempo di $(x_i, y_i) \in M \times N$ è descritta da un **prodotto obliquo**

$$\begin{aligned}x_{i+1} &= f(x_i, y_i) = f_{y_i}(x_i) \\ y_{i+1} &= g(y_i)\end{aligned}$$

- Le generalizzazioni possibili di Takens sono diverse se conosciamo (N, g) oppure no.

Il primo adattamento

- Se non conosciamo (N, g) potremmo pensare di applicare Takens prendendo il sistema $M \times N$ per intero.
- Il problema è che vorremmo usare un osservabile che dipende solo dallo stato di M .

Teorema (Takens forzato - Stark 1999)

Siano M^m e N^n varietà compatte ($m \geq 1$) e siano $r \geq 1$, $d \geq 2(m + n) + 1$.

Supponiamo che $g \in \text{Diff}^r(N)$ sia t.c. le orbite periodiche di periodo $< 2d$ siano isolate; e che Tg , calcolato nei punti di tali orbite, abbia sempre autovalori tutti distinti.

Allora esiste un aperto denso di

$$(f, \phi) \in \text{Diff}^r(M \times N, M) \times \mathcal{C}^r(M, \mathbb{R})$$

tale che la delay map $\Phi_{(f,g),\phi}$ è un embedding di $M \times N$.

La ricostruzione nel caso forzato

- Tramite Takens forzato, ogni mappa f_y risulta coniugata a $F_y = \Phi_{(f,g),\phi,g(y)} \circ f_y \circ \Phi_{(f,g),\phi,y}^{-1}$. Conoscere l'evoluzione temporale di y è quindi necessario.
- Se (N, g) è noto a priori, potremmo abbassare d richiedendo solo che gli $\Phi_{(f,g),\phi,y}$ siano embedding.

Teorema (degli embedding fibrati - Stark 1999)

Siano M^m e N^n varietà compatte ($m \geq 1$) e siano $r \geq 1$, $d \geq 2m + 1$. Supponiamo che $g \in \text{Diff}^r(N)$ sia t.c. le orbite periodiche di periodo $< d$ abbiano misura di Lebesgue nulla in N .

*Allora esiste un insieme **residuo** di*

$$(f, \phi) \in \text{Diff}^r(M \times N, M) \times \mathcal{C}^r(M, \mathbb{R})$$

tale che la delay map $\Phi_{(f,g),\phi,y}$ è un embedding di M al variare di y in un insieme aperto, denso e di misura piena in N .

Alcune osservazioni

- La tesi di questa versione è più debole delle precedenti.
- Le ipotesi tecniche su g per Takens forzato sono necessarie.
- Spesso, quando si ha una forzatura periodica, si campiona ϕ con una frequenza pari a quella della forzatura, quindi $g = id_N$. Però in tal caso il sistema M risulta non forzato.
- Si costruiscono esempi di g per cui la tesi non vale per *tutti* gli y per un insieme aperto di (f, ϕ) .

Dal caso forzato a quello stocastico

I sistemi stocastici di nostro interesse si possono modellizzare come una generalizzazione di quelli forzati: l'evoluzione di M è data da una f_{ω^0} con $\omega^0 \in N$ scelta casualmente ad ogni passo.

- Sia $\Sigma = N^{\mathbb{Z}}$ e sia σ lo **shift** su Σ : $[\sigma(\omega)]^i = \omega^{i+1}$.
- Il sistema $M \times \Sigma$ si evolve ancora secondo un prodotto obliquo

$$\begin{aligned} x_{i+1} &= f(x_i, \omega_i) = f_{\omega_i}(x_i) \\ \omega_{i+1} &= \sigma(\omega_i) \end{aligned}$$

- La delay map diventa

$$\Phi_{f,\phi,\omega} : x \mapsto (\phi(x), \phi(f_{\omega^0}x), \phi(f_{\omega^0\omega^1}x) \dots, \phi(f_{\omega^0\dots\omega^{d-2}}x))$$

- Ci troviamo in un caso poco dissimile da quello forzato finito-dimensionale.

Un'ambientazione più probabilistica

- Su N si può porre una misura di probabilità μ , e su Σ quella prodotto.
- Σ può anche essere un sottoinsieme σ -invariante di $N^{\mathbb{Z}}$ dotato di una misura μ_{Σ} anch'essa σ -invariante.

Teorema (Takens per processi stocastici - Stark et al. 2003)

Siano M^m e N^n varietà compatte ($m \geq 1$) e siano $r \geq 1$, $d \geq 2m + 1$. Sia μ_{Σ} una misura di probabilità σ -invariante su Σ t.c. la misura marginale μ_{d-1} su N^{d-1} sia assolutamente continua rispetto alla misura di Lebesgue.

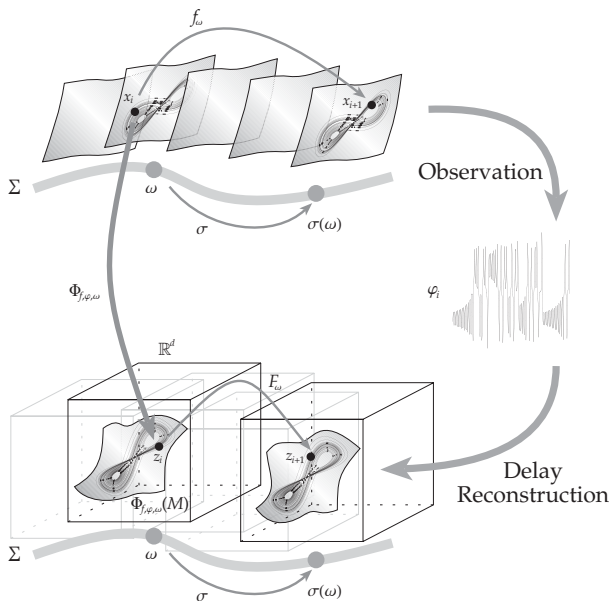
Allora esiste un insieme residuo di

$$(f, \phi) \in \text{Diff}^r(M \times N, M) \times C^r(M, \mathbb{R})$$

tale che la delay map $\Phi_{f, \phi, \omega}$ è un embedding di M per μ_{Σ} -quasi ogni ω .

I passi della dimostrazione

- $\Phi_{f,\phi} : M \times N^{d-1} \rightarrow \mathbb{R}^d$ è tale che $\tilde{T}\Phi_{f,\phi} \pitchfork$ alla sezione nulla di $T\mathbb{R}^d$ e $\Phi_{f,\phi} \times \Phi_{f,\phi} \pitchfork \Delta \subset \mathbb{R}^d \times \mathbb{R}^d$ per un insieme residuo di (f, ϕ) .
- Per il **teorema di trasversalità parametrica** questo è vero per quasi tutte le $\Phi_{f,\phi,\omega}$.
- Per questioni di dimensione, trasversalità=non intersezione.
- Il problema principale nella dimostrazione sono i punti che prendono due o più coordinate uguali.



Alcune osservazioni

- Per $n = 0$, si ha un embedding per *ogni* ω .
- Il teorema è applicabile al caso in cui c'è soltanto un rumore stocastico (anche nell'osservabile).
- È applicabile anche al caso in cui si fanno campionamenti di un sistema deterministico a intervalli irregolari.
- Il teorema non dà alcuna indicazione sulla ricostruzione di ω , necessaria per ricostruire M .
- Non sembra possibile caratterizzare l'insieme delle (f, ϕ) “buone”.

Serie temporali vettoriali

Se abbiamo più osservabili ϕ^1, \dots, ϕ^m , ha senso pensare ad una delay map del tipo

$$\begin{aligned} \mathbf{x} \mapsto & (\phi^1(\mathbf{x}), \phi^1(f^{(\tau_1)}\mathbf{x}), \dots, \phi^1(f^{((d_1-1)\tau_1)}\mathbf{x}), \\ & \vdots \\ & \phi^m(\mathbf{x}), \phi^m(f^{(\tau_m)}\mathbf{x}), \dots, \phi^m(f^{((d_m-1)\tau_m)}\mathbf{x})) \end{aligned}$$

- Non c'è alcun ostacolo teorico ad applicare Takens.
- Data una serie temporale $(\phi_i = (\phi_i^1, \dots, \phi_i^m))_{i=1}^N$, siano \mathbf{V}_i i relativi vettori di ricostruzione ($i = J_0, \dots, N$).
- Come nel caso standard, avremo $\mathbf{V}_{i+1} = F(\mathbf{V}_i)$; o equivalentemente, si hanno delle mappe F^k tali che $\phi_{i+1}^k = F^k(\mathbf{V}_i)$.

Scelta dei parametri

- Risulta meglio scegliere i τ_j e i d_j indipendentemente per ogni F^k .
- L'idea per i d_j è fare in modo che le F^k risultino continue.
- Sia $\eta(i)$ l'indice per cui $\|\mathbf{V}_i - \mathbf{V}_{\eta(i)}\|$ è minimo.
- L'errore medio nell'approssimazione grezza di F^k è

$$E^k(d_1, \dots, d_m) = \frac{1}{N-J_0+1} \sum_{i=J_0}^N |\phi_{i+1}^k - \phi_{\eta(i)+1}^k|.$$
- Ci si aspetta che E^k presenti un punto di minimo.
- La scelta più opportuna è quindi $(d_1, \dots, d_m) = \operatorname{argmin} E^k$.
- Le F^k vengono approssimate in modo **localmente lineare**.
- Se vogliamo approssimare le iterate di F , bisogna ridefinire l'errore di conseguenza.

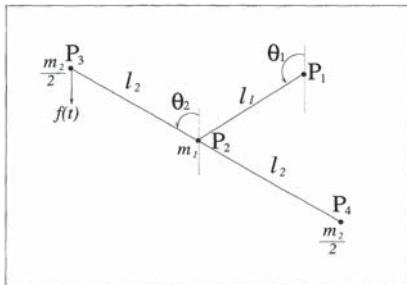
Utilizzi dell'approccio multivariato

- Sulla **predizione** di una serie temporale, l'approccio multivariato lavora molto meglio di quello univariato.
- Si possono rintracciare le **relazioni funzionali** esistenti fra più serie: si cerca di fittare una relazione del tipo

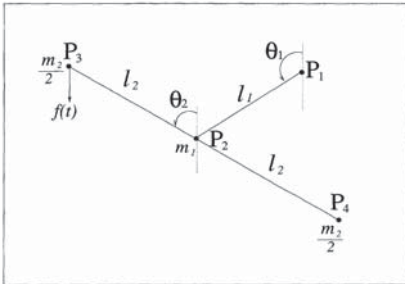
$$\begin{aligned} \phi_i^l &= G(\phi_{i-t_{11}}^{k_1}, \dots, \phi_{i-t_{1d_1}}^{k_1}, \\ &\quad \vdots \\ &\quad \phi_{i-t_{m1}}^{k_m}, \dots, \phi_{i-t_{md_m}}^{k_m}) \end{aligned}$$

che può essere **predittiva** o meno.

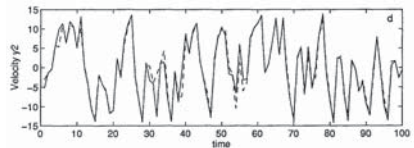
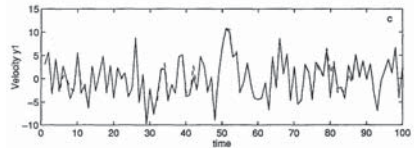
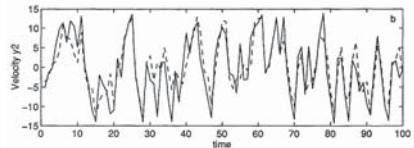
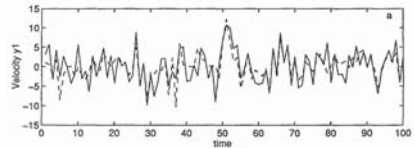
La figura illustra il cosiddetto
doppio rotore calciato



La figura illustra il cosiddetto
doppio rotore calciato



A destra sono i risultati della
predizione nel caso univariato e
bivariato.



Connettività funzionale fra diverse aree cerebrali

- La **Functional Magnetic Resonance Imaging** osserva l'attività dei neuroni cerebrali nel tempo, suddivisi in **voxel**.
- È un modo per indagare su quali aree rispondono a determinati stimoli.
- Altra domanda importante è la **connettività funzionale**: la “correlazione temporale fra eventi neurofisiologici lontani nello spazio”.
- Usualmente la correlazione fra X e Y è definita da
$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$
- Con tale definizione, però, si rintracciano solo i legami a livello lineare.

Bivariate Nonlinear Connectivity Index

- Consideriamo due serie temporali (ϕ_i) e (ψ_i) provenienti da voxel diversi.
- Siano d_1^u, d_2^u le rispettive dimensioni ottimali di embedding univariato.
- Sia invece d^b la dimensione totale di embedding bivariato.
- Quando si ha connettività, ci si aspetta che $d^b < d_1^u + d_2^u$.
- Si definisce quindi

$$BNC = 1 - \frac{|d^b - d_1^u| + |d^b - d_2^u|}{d_1^u + d_2^u}.$$

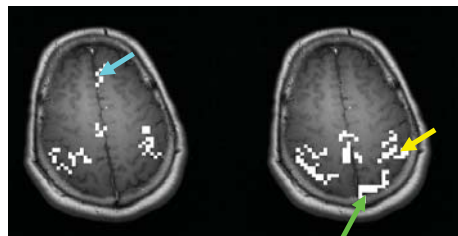
- Il **metodo** δ - ϵ viene utilizzato come test per assicurarsi un legame deterministico piuttosto che dovuto a rumore: a livello quantitativo viene utilizzato

$$S(r) = |\epsilon(r) - \epsilon^*(r)| / \sigma^*(r)$$

FC in stato di riposo

| Network | | Linear | Nonlinear | Granger Causality | |
|---------|--------------------------|-------------------|--------------------|----------------------|-------|
| | | Connectivity (LC) | Connectivity (BNC) | Direction | DGC |
| Sub-1 | LM \leftrightarrow SMA | 0.76 | 0.52 | LM \rightarrow SMA | -0.52 |
| | LM \leftrightarrow RM | 0.73 | 0.89 | LM \rightarrow RM | -0.78 |
| | LM \leftrightarrow F | 0.22 | 0.15 | LM \rightarrow F | 0.20 |
| Sub-2 | LM \leftrightarrow SMA | 0.49 | 0.63 | LM \rightarrow SMA | -0.56 |
| | LM \leftrightarrow RM | 0.57 | 0.76 | LM \rightarrow RM | -0.65 |
| | LM \leftrightarrow F | 0.45 | 0.54 | LM \rightarrow F | 0.15 |

- I risultati dell'analisi nonlineare sembrano precisare, senza contraddire, quelli dell'analisi lineare.
- Il risultato ottenuto tramite analisi nonlineare appare più solido.

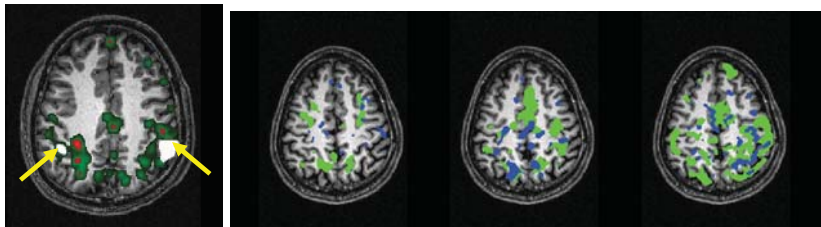


Delta-Epsilon

Correlation

Il metodo nonlineare trova una zona connessa in più, mentre elimina una connessione spuria trovata tramite correlazione.

FC mentre si tamburellano le dita



- Le figure mostrano i voxel che presentano una FC significativa con la regione di riferimento in bianco.
- I voxel connessi aumentano con il trascorrere del tempo.
- L'analisi nonlineare rileva molti più voxel connessi di quella lineare.

Grazie per l'attenzione.

Riferimenti bibliografici



Andrea Marchesini

Ricostruzione di attrattori a partire da serie temporali: il teorema di Takens
Tesi di Laurea triennale, A.A. 2009-2010



J. Stark

Delay Embeddings for Forced Systems. I. Deterministic Forcing
J. Nonlinear Sci. Vol. 9: pp. 255-332, 1999 Springer-Verlag



J. Stark, D.S. Broomhead, M.E. Davies, J. Huke

Delay Embeddings for Forced Systems. II. Stochastic Forcing
J. Nonlinear Sci. Vol. 13: pp. 519-577, 2003 Springer-Verlag



Liangyue Cao, Alistair Mees, Kevin Judd

Dynamics from multivariate time series
Physica D 121, 1998, pp. 75-88



Gopikrishna Deshpande, Stephen LaConte, Scott Peltier, Xiaoping Hu

Connectivity Analysis of Human Functional MRI Data: From Linear to Nonlinear and Static to Dynamic
MIAR 2006, LNCS 4091, pp. 17-24, 2006 Springer-Verlag



Andrew M. Fraser, Harry L. Swinney

Independent coordinates for strange attractors from mutual information
Physical Review A, Vol. 33, No. 2, pp. 1134-1140, 1986 The American Physical Society