Analyse von Abhängigkeiten zwischen Zeitreihen

Andreas Schmitz und Jochen Arnhold

Theoretische Physik, Universität Wuppertal & NIC, FZ Jülich

I. Einleitung

Die Frage nach Abhängigkeiten zwischen gemessenen Signalen stellt sich in vielen Bereichen. Sie kann zum einen praktischer Natur sein, wenn es darum geht, große Datenmengen zu reduzieren und die Analyse redundanter Informationen zu verhindern. Auf der anderen Seite besteht oft auch großes Interesse an den Abhängigkeiten zwischen physikalisch *unterschiedlichen* Größen eines Systems. Schließlich bilden Abhängigkeitsmaße ein wichtiges Werkzeug bei der Analyse gekoppelter Systeme und deren Synchronisation.

Verschiedene Maße für Abhängigkeit werden auf ihre Leistungsfähigkeit hin untersucht. Dabei wird neben linearen und informationstheoretischen Größen ein neuer Ansatz vorgestellt, der als entscheidendes zusätzliches Merkmal die Asymmetrie aufweist. Dies kann es unter Umständen ermöglichen, auch die *Richtung* der Kopplung zwischen zwei Systemen aus den Daten zu extrahieren.

II. Abhängigkeitsmaße

Lineare Korrelationen

Die erste Größe ist die aus der linearen Zeitreihenanalyse stammende Kreuzkorrelation

$$C_{\text{lin}}(X,Y) = \frac{1}{N}\sum_{i=1}^{N} x_i y_i \qquad (1)$$

Sind die Zeitreihen $\{x_i\}$ und $\{y_i\}$ $(i = 1,\ldots,N)$ auf Varianz $\sigma^2 = 1$ und den Mittelwert $\bar{x} = \bar{y} = 0$ skaliert, liegt C_{lin} im Intervall [0,1]. Es ist zu beachten, daß sich aus einer Korrelation $C_{\text{lin}} \neq 0$ eine Abhängigkeit zwischen den Variablen X und Y schließen läßt, der Umkehrschluß aber nicht gilt.

Gegenseitige Information

Aus der Informationtheorie entliehen ist die gegenseitige Information

$$M(X,Y) = I(X) + I(Y) - I(X,Y) \qquad (2)$$

Dabei ist $I(\cdot)$ die Entropie (Information) und $I(X,Y)$ die zusammengesetzte Entropie der Variablen X und Y. Die Entropien lassen sich über eine Paritionierung in B Boxen und Abschätzung der Boxwahrscheinlichkeiten p_i berechnen

$$I(X) = -\sum_{i=1}^{B} p_i^{(x)} \log p_i^{(x)} \qquad (3)$$

Alternativ kann die verallgemeinerte Entropie I_q $(I = I_1)$ über die verallgemeinerte Korrelationssumme C_q bestimmt werden

$$I_q(\cdot,\delta) \approx -\log C_q(\cdot, r = 2\delta) \qquad (4)$$

wobei δ die Boxgröße und r der Nachbarschaftsradius ist. Die Verwendung von C_2 hat den Vorteil, daß es am schnellsten berechnet werden kann und keine *finite-sample*-Korrektur benötigt.

Neue Maße

Ein neuer Ansatz gegenseitige Abhängigkeit zu bestimmen wurde in Ref.[1] vorgestellt. Dabei werden zunächst für jeden Punkt $\mathbf{x}_i$ (im Phasenraum bzw. Einbettungsraum) die k nächsten Nachbarn mit Indizes $r_i(j)$ mit $j = 1,\ldots,k$ gesucht. Der quadrierte, mittlere Abstand ist gegeben durch

$$R_i^{(k)}(X) = \frac{1}{k}\sum_{j=1}^{k}(\mathbf{x}_i - \mathbf{x}_{r_i(j)})^2 \qquad (5)$$

Analog läßt sich $R_i^{(k)}(Y)$ bestimmen, wobei $s_i(j)$ die Indizes der k nächsten Nachbarn von $\mathbf{y}_i$ seien. Weiterhin läßt sich der Y*-bedingte* Abstand

$$R_i^{(k)}(X|Y) = \frac{1}{k}\sum_{j=1}^{k}(\mathbf{x}_i - \mathbf{x}_{s_i(j)})^2 \qquad (6)$$

definieren, welcher sich nur in den Indizes von $R_i^{(k)}(X)$ unterscheidet. Für völlig unabhängige Systems ist $R_i^{(k)}(X|Y) \gg R_i^{(k)}(X)$, während stark abhängige Systeme $R_i^{(k)}(X|Y) \approx R_i^{(k)}(X)$ liefern. Daraus kann ein Abhängigkeitsmaß über

$$S^{(k)}(X|Y) = \frac{1}{N}\sum_{i=1}^{N}\frac{R_i^{(k)}(X)}{R_i^{(k)}(X|Y)} \qquad (7)$$

definiert werden, welches Werte von 0 (für unabhängige Systeme) bis 1 annehmen kann. Eine ganz ähnliche Größe ist

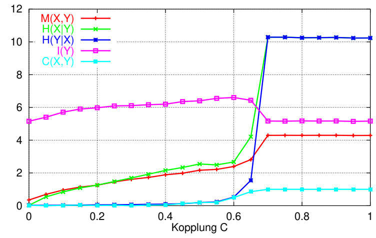
$$H^{(k)}(X|Y) = \frac{1}{N}\sum_{i=1}^{N}\log\frac{R_i^{(N-1)}(X)}{R_i^{(k)}(X|Y)} \qquad (8)$$

welche sich im wesentlichen durch die Mittelung über *alle* $N-1$ Nachbarn im Zähler unterscheidet. Durch diese Änderung ist H robuster und systematischer als S, reagiert aber auch sensitiver auf Ausreißer.

III. Vergleich der Maße

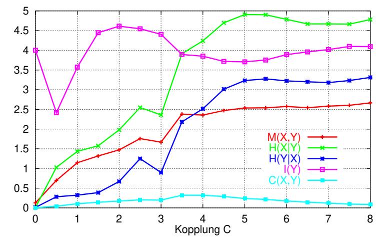
Gekoppelte, chaotische Systeme

Als erstes Beispiel dienen zwei einseitig gekoppelte Hénon–Abbildungen, wobei X das treibende und Y das getriebene System ist[2]. Die unterschiedlichen Abhängigkeitsmaße werden in Abhängigkeit von der Kopplungsstärke C betrachtet. Für $C > 0.7$ synchronisieren die Systeme identisch.



Sowohl die gegenseitige Information M als auch H zeigen einen deutlichen Anstieg bei $C = 0.7$

Im zweiten Beispiel treibt ein Rössler–System X ein Lorenz–System Y. Für $C \approx 2.5$ und $C > 3$ synchronisieren die Systeme.



In beiden Fällen zeigen M und H deutliche Zeichen für das Eintreten der Synchronisation, im Gegensatz zur linearen Kreuzkorrelation C_{lin} im zweiten Beispiel. Außerdem gilt in diesen Beispielen stets

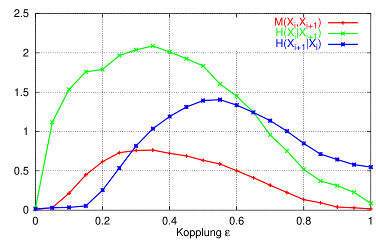
$$H(X|Y) \geq H(Y|X) \qquad (9)$$

woran die Richtung der Kopplung abgelesen werden kann.

Gekoppelte Abbildungen auf dem Gitter

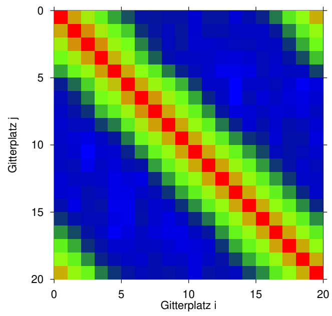
Die gezeigte Asymmetrie von H könnte sich aber auch durch einen Unterschied von Eigenschaften auf den *einzelnen* Zeitreihen ergeben. So gilt z.B. in beiden Beispielen auch stets $I(Y) \geq I(X)$. Um diese Alternative auszuschließen wird ein eindimensionales, periodisches Gitter mit einseitig gekoppelten Zelt–Abbildungen f betrachtet:

$$x_i^{t+1} = (1-\varepsilon)f(x_i^t) + \varepsilon f(x_{i-1}^t) \qquad (10)$$

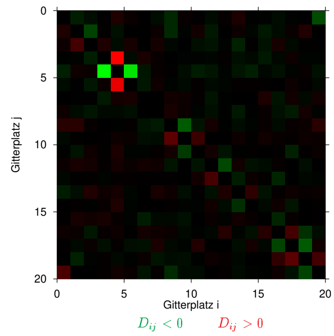


Die Asymmetrie von H bleibt auch in diesem Fall bestehen, während univariate Größen wie z.B. die Entropie auf jedem Gitterplatz gleich sind.

Im folgenden Beispiel sind die Abbildungen auf dem Gitter in beide Richtungen gekoppelt. Dabei fällt $H(X_i|X_j)$ mit steigendem Abstand $|i-j|$ zwischen den Gitterpunkten i und j ab (Farbskala logarithmisch):



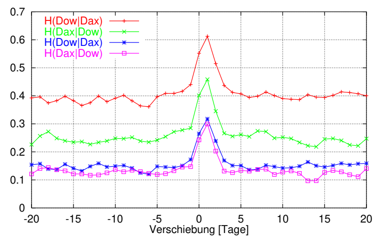
Wird auf einen Gitterplatz eine Zelt–Abbildung mit anderer Steigung gesetzt, ist dies deutlich in $D_{ij} = H(X_i|X_j) - H(X_j|X_i)$ zu sehen:



IV. Anwendungen

Finanzdaten

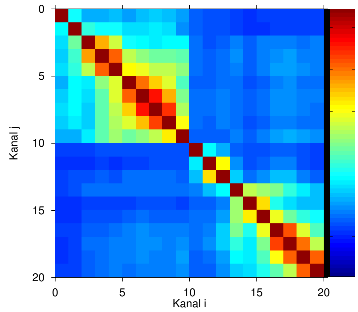
Im folgenden werden gegenseitige Abhängigkeiten der Börsenindizes Dow Jones und DAX mit Hilfe von H untersucht. Die verwendeten Zeitreihen bestehen aus den Logarithmusdifferenzen der Notierungen.



Für die beiden unteren Kurven wurden alle Crashs herausgenommen, wodurch sich eine kleinere gegenseitige Abhängigkeit ergibt. In beiden Fällen zeigen sich aber Anzeichen dafür, daß der Verlauf des DAX vom gestrigen Dow Jones abhängt.

EEG–Daten

Sehr interessante Anwendungen ergeben sich in der Analyse von EEG–Daten von Patienten mit epileptischen Anfällen. Die Daten wurden uns freundlicherweise von der Klinik für Epileptologie in Bonn zur Verfügung gestellt. Im folgenden ist S_{ij} für 20 Elektroden (10 auf jeder Seite) eines Patienten berechnet worden:



Zunächst ist deutlich eine Trennung zwischen rechter und linker Hemisphäre zu erkennen. Außerdem zeigt sich ein erhöhter Wert von S_{ij} um Kanal 8, was auf stärkere Synchronisation und einen möglichen Fokus hindeutet. In der Tat wurde der Patient an dieser Stelle operiert und ist nun anfallsfrei.

V. Zusammenfassung

Erste Beispiele und Anwendungen zeigen, daß die neuen Maße S und H ein nützliches Werkzeug bei der Analyse von Abhängigkeiten und Synchronisation sind. Von entscheidendem Vorteil gegenüber anderen Größen ist dabei deren Asymmetrie, wodurch auch Kopplungsrichtungen abgelesen werden können.

Literatur

[1] J. Arnhold, P. Grassberger, K. Lehnertz und C.E. Elger, *A Robust Method for Detecting Interdependences: Application to Intracranially Recorded EEG*, to be published in Physica D (1999)

[2] R. Quian Quiroga, J. Arnhold und P. Grassberger, *Can Driver-Response Relationships be Learned from Synchronization Patterns?*, submitted to Phys. Rev. E (1999)