Vorhersagende Information

nach Bialek, Nemenman und Tishby

(Predicability, Complexity and Learning; 2001)

Bild: wikipedia.org GPL

ein Vortrag von Felix Dollinger

Gliederung

1) Einleitung
2) Beispiel
3) Definition *Vorhersagende Information*
4) Berechnung und...
5) ...Deutung der vorhersagenden Information
6) Zusammenfassung

Einleitung (I)

- Vorhersagen waren für die Menschen schon immer von Bedeutung.
- Gabe der Antizipation ist wichtig für z.B. für:
    - SportlerInnen
    - BörsenmaklerInnen
    - PolitikerInnen
- Reaktionen können nur von vorhersagender Information beeinflusst werden.

Einleitung (II)

Bialek et al.:

„Nonpredictive Information is useless to the organism“

?


Vorhersagende Information ist wichtig für den Organismus.
Photo: eigenes Werk

Einleitung (III)

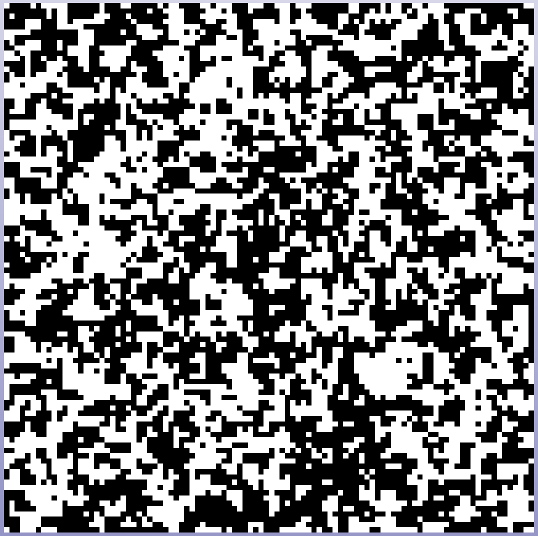
- Nur ein Teil der aufgenommenen Information ist vorhersagend
- Bsp: Zufallszahlenreihe sagt nichts über ihre „Zukunft“ aus.

=> Interesse daran, die vorhersagende Information zu isolieren.

- Quantifizierung durch die Größe I_{pred} [bit]

Beispiel (I)

- Ising Spins: Ernst Ising 1925
- Modell für Magnetismus aus der statistischen Physik



Animation: wikipedia.org GPL

- Hier: **eindimensionale** Kette aus *AN/AUS*-Zuständen

Beispiel (II)

- Betrachtung dreier „Rechengenauigkeiten“:
  1) Alle Spin-Teilchen unabhängig --> Zufallsreihe
  2) nur Beeinflussung durch direkten Nachbar
  3) Abstandsabhängige Fernbeeinflussung

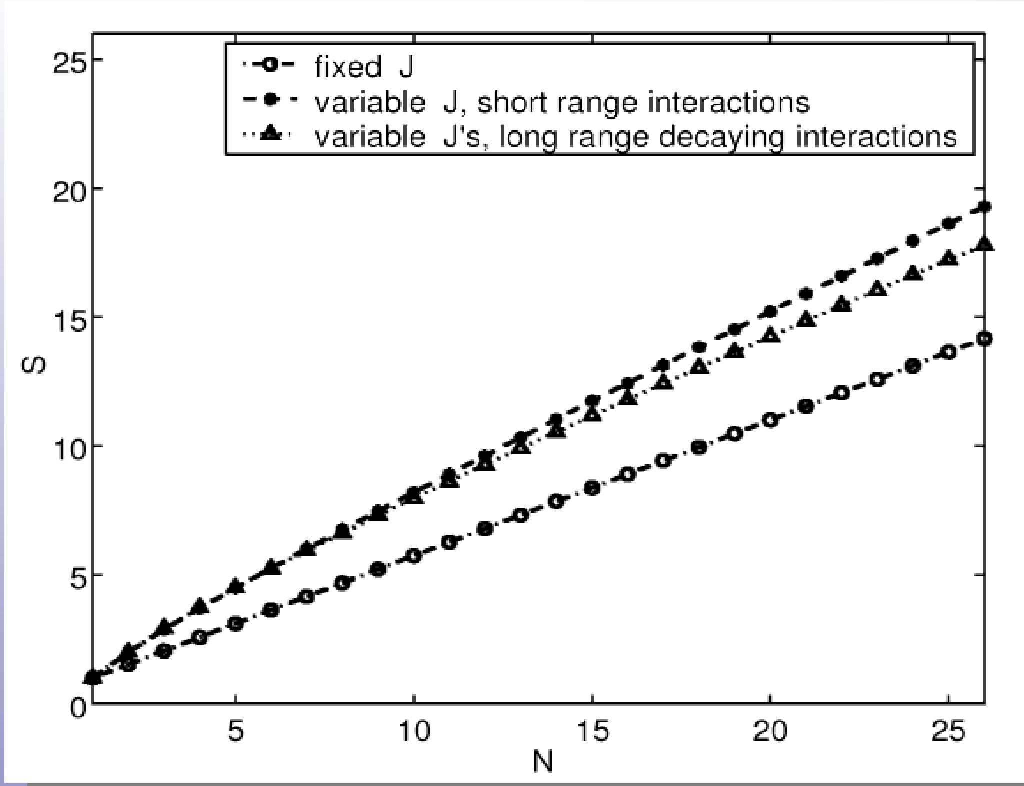
- Berechnung der Entropie H(N) von „Wörtern“ aus der Spinreihe.

$$H(N) = -\sum_{k=0}^{2^N-1} P_N(W_k)\log_2(P_N(W_k))$$

$$\text{mit } W_k\text{: } k\text{-tes Wort der Länge } N$$

Beispiel (III)

Plot der Entropie:



Plot: Bialek et al. 2001 S=H

jeweils *annähernd* **extensive** (lineare) Entropiekurve

Beispiel (IV)

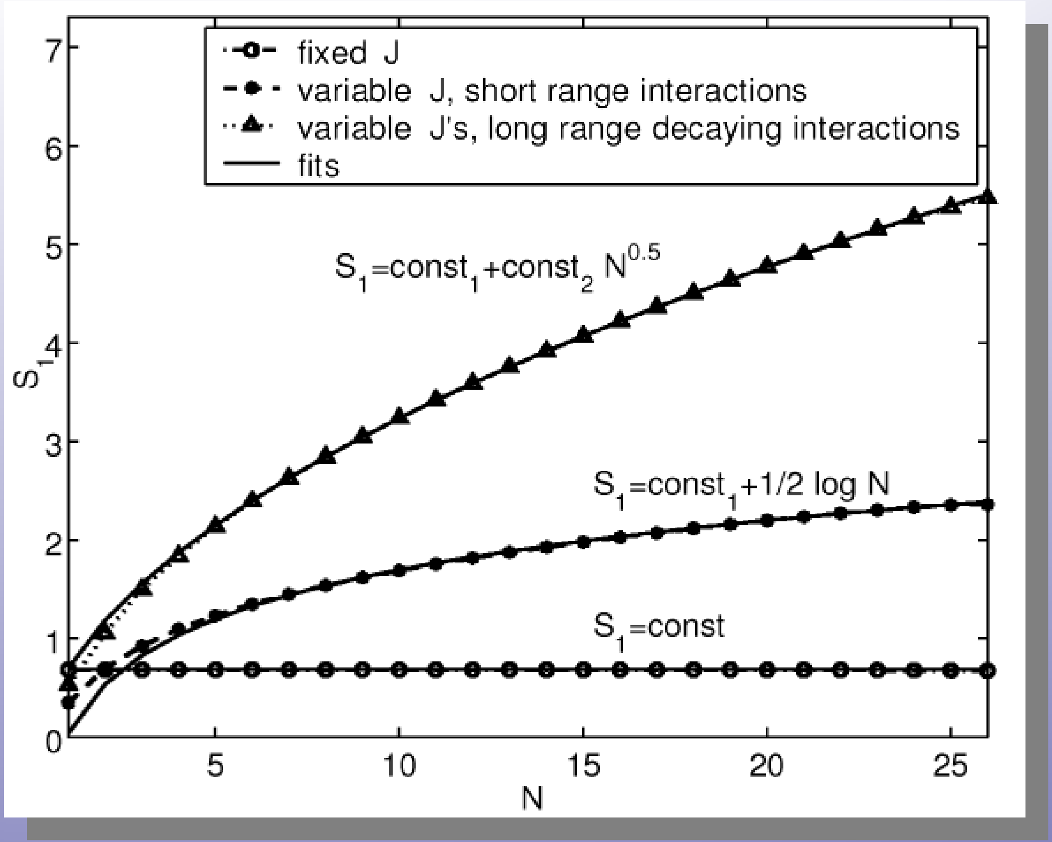
Subtraktion des linearen Anteils der Entropie:

$$H_1(N) = H(N) - \varsigma_0 * N$$

1) konstanter Rest (H_1)

2) logarithmischer Rest

3) Wurzelfunktion als Rest



Plot: Bialek et al. 2001 S=H

Beispiel (V)

Durch den Zusammenhang aus früheren und späteren Elementen ergibt sich ein Zuwachs an Entropie (~Informationsgehalt) des einzelnen Worts über das ganze System.

Wissen über frühere Teile des Systems

→ Wissen über spätere Teile des Systems

Wie viel?

Definition

Vorhersagende Information

1) Die Information, die die Vergangenheit über die Zukunft enthält.

oder

2) Die Information, die wir zu viel aufwenden, wenn wir Vergangenheit und Zukunft jeweils als unabhängig betrachten.

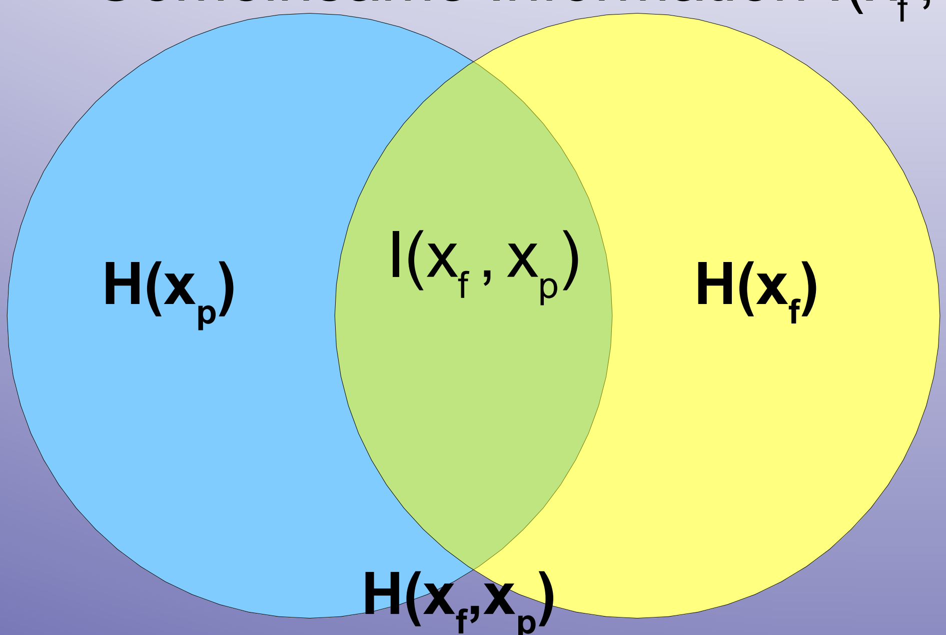
Berechnung (I)

1) Die Information, die die Vergangenheit über die Zukunft enthält.

<u>Def.:</u> Daten in einem Zeitintervall:

$$x_p := x(t),\ -T<t<0 \quad x_f := x(t),\ 0<t<T'$$

Gemeinsame Information $I(x_f, x_p)$:



$$\boldsymbol{I}(\boldsymbol{x_f},\boldsymbol{x_p}) = \boldsymbol{H}(\boldsymbol{x_p}) + \boldsymbol{H}(\boldsymbol{x_f}) - \boldsymbol{H}(\boldsymbol{x_f},\boldsymbol{x_p})$$
$$= \boldsymbol{I_{pred}}(\boldsymbol{T},\boldsymbol{T'})$$

Sommerakademie Olang

Berechnung (II)

1) Die Information, die die Vergangenheit über die Zukunft enthält.

$$I_{pred}(T,T')=H(x_p)+H(x_f)-H(x_f,x_p)$$

$$=-\langle ld(P(x_p))\rangle-\langle ld(P(x_f))\rangle+\langle ld(P(x_f,x_p))\rangle$$

$$=-\langle ld\frac{P(x_f,x_p)}{P(x_f)*P(x_p)}\rangle$$

siehe „Bayessche Graphen“

$$=-\langle ld\frac{P(x_f|x_p)}{P(x_f)}\rangle$$

$$=H(T)+H(T')-H(T+T')$$

da wir uns für Eigenschaften der Information interessieren, die fortbestehen, können wir zeitliche Veränderungen der Daten x vernachlässigen.

=> Abhängigkeit der Entropie von der Dauer der Beobachtung T.

Berechnung (III)

2) Die Information, die wir zu viel aufwenden, wenn wir Vergangenheit und Zukunft jeweils als unabhängig betrachten.

$$I_{pred}(T, T') = D_{KL}(P(x_f, x_p) \| P(x_f) * P(x_p))$$

$$= \sum_x P(x_f, x_p)\, ld \frac{P(x_f, x_p)}{P(x_f) * P(x_p)}$$

$$= -\left\langle ld \frac{P(x_f, x_p)}{P(x_f) * P(x_p)} \right\rangle$$

$$= -\left\langle ld \frac{P(x_f | x_p)}{P(x_f)} \right\rangle$$

siehe „Relative Entropie“

siehe voriger Slide

All diese Formeln sind symmetrisch: x_p und x_f sind vertauschbar.

Deutung (I)

Wenn x_p und x_f vertauschbar sind gilt:

Es ist genauso schwer, auf die Zukunft zu schließen, wie auf die Vergangenheit.

-> Interessant in Kriminalistik, Archäologie,...

?

42

?

Diese Quizshow ist seit 1964 erfolgreich.

?

?

Deutung (II)

kurz zurück zum Beispiel

$$H_1(N) = H(N) - \varsigma_0 * N \longrightarrow H(N) = H_1(N) + \varsigma_0 * N$$

$$I_{pred}(T, T') = H(T) + H(T') - H(T + T')$$

$$\Rightarrow \quad I_{pred}(T, T') = H_1(T) + \cancel{\varsigma_0 * T} + H_1(T') + \cancel{\varsigma_0 * T'}$$
$$- H_1(T + T') - \cancel{\varsigma_0 * T} - \cancel{\varsigma_0 * T'}$$

=> Die vorhersagende Information gibt die Abweichung von der Extensivität der Entropie an.

Deutung (II)

Aus *Relative Entropie* von Lennart Hilbert:

„Die Entropie entspricht der Länge der komprimierten Sequenz“

=> eigentlich ist die Entropie eine extensive Größe. Hier nur für große T:

$$\lim_{T\to\infty}\left(\frac{H(T)}{T}\right)=\varsigma_0$$

Dann muss wegen $H(T)=H_1(T)+\varsigma_0 * T$ gelten:

$$\lim_{T\to\infty}\left(\frac{H_1(T)}{T}\right)=0$$

Deutung (III)

$$\lim_{T\to\infty}\left(\frac{H(T)}{T}\right)=\varsigma_0 \qquad \lim_{T\to\infty}\left(\frac{H_1(T)}{T}\right)=0$$

- Aussage aus einer Beobachtung der Zeit T über die gesamte Zukunft:

$$\lim_{T\to\infty}(I_{pred}(T,T'))$$
$$=\lim_{T\to\infty}(H_1(T)+H_1(T')-H_1(T+T'))$$
$$=H_1(T)=I_{pred}(T)$$

- Die über eine Zeit T gesammelte Information H(T) wächst schneller als die vorhersagende Information $I_{pred}(T)=H_1(T)$

$$\lim_{T\to\infty}\left(\frac{I_{pred}(T)}{H(T)}\right)=0$$

Deutung (IV)

•Die meisten Beobachten sind irrelevant um Vorhersagen zu treffen.

•Je länger man ein bestimmtes System beobachtet, desto ineffektiver wird das.

<=> Die Lernkurve $\Lambda(\boldsymbol{T}) = \frac{\partial \boldsymbol{I}_{pred}(\boldsymbol{T})}{\partial \boldsymbol{T}}$ geht für große T gegen Null.

Deutung (V)

Komplexität

- Die vorhersagende Information einer Zufallsreihe ist 0.
- Mit zunehmender Interdependenz der Elemente steigt die vorhersagende Information.

=> Sie eignet sich als ein Maß der Komplexität.

Zusammenfassung

Die vorhersagende Information I_{pred}

- ist die gemeinsame Information von Vergangenheit und Gegenwart.
- ist Grundlage aller (Re)Aktionen.
- ist ein kleiner Teil der gesamten Information.
- sagt nichts darüber **welcher** Teil wichtig ist.
- wächst sublinear mit der Beobachtungszeit.
- misst die Komplexität eines Systems.

Danke für eure Aufmerksamkeit!

Quelle: Predictability, Complexity and Learning; Bialek et al. 2001