



Das Unvorhersehbare vorhersagen:
Representation Learning von Zeitreihen

Marisa Mohr

Hamburg, 12. Juni 2019



Marisa Mohr

M.Sc. Mathematik

 @MarsaMo1

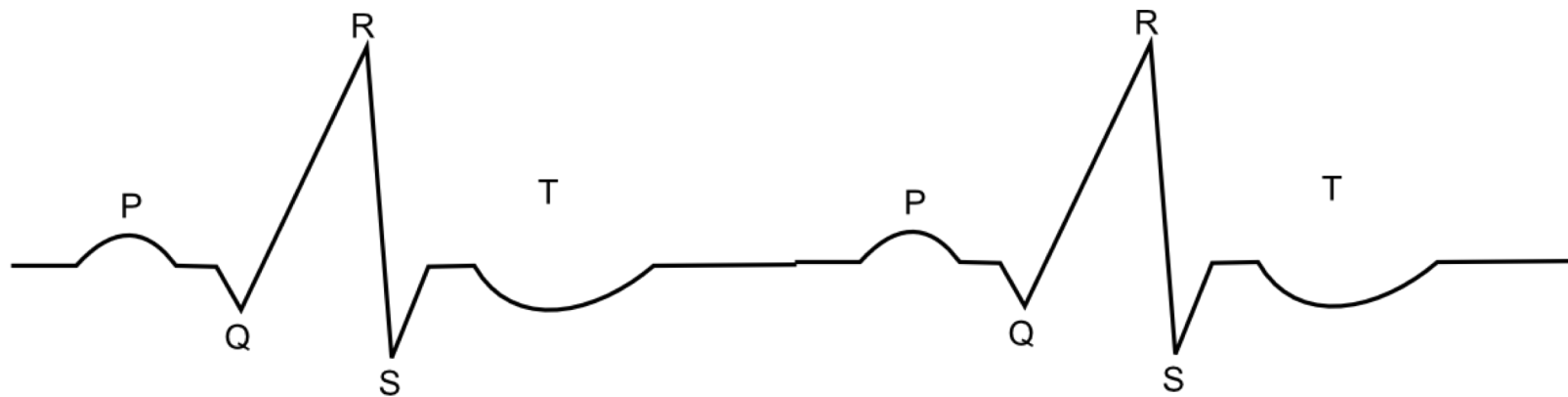
Machine Learning Engineer @inovex

- Machine Perception & Artificial Intelligence
- IoT und Predictive Maintenance

Doktorandin @Universität zu Lübeck

- Institut für Informationssysteme, Prof. Dr. Ralf Möller
- “Time-Series Representation Learning”

Dynamisches Auf und Ab



The image features two wireframe heads, one on the left and one on the right, rendered in a light blue color. Each head is composed of a complex network of lines and nodes, resembling a neural network or a digital brain. The background is a dark blue gradient with horizontal light blue lines. A dark blue rectangular box is centered in the middle of the image, containing the text "High Dimensionality" in white.

High Dimensionality



Low Sample Size

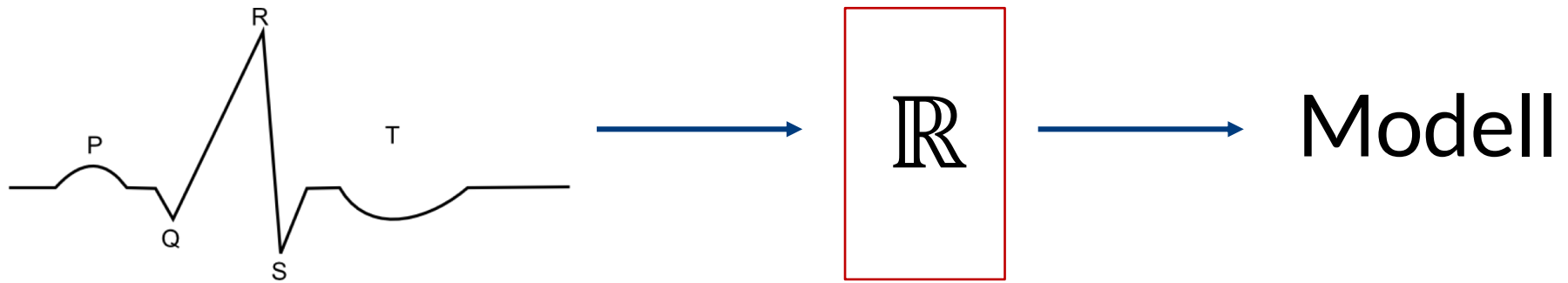


Interpretierbarkeit

Die innere Natur
des erzeugenden Systems
einer Zeitreihe verstehen.

Representation Learning

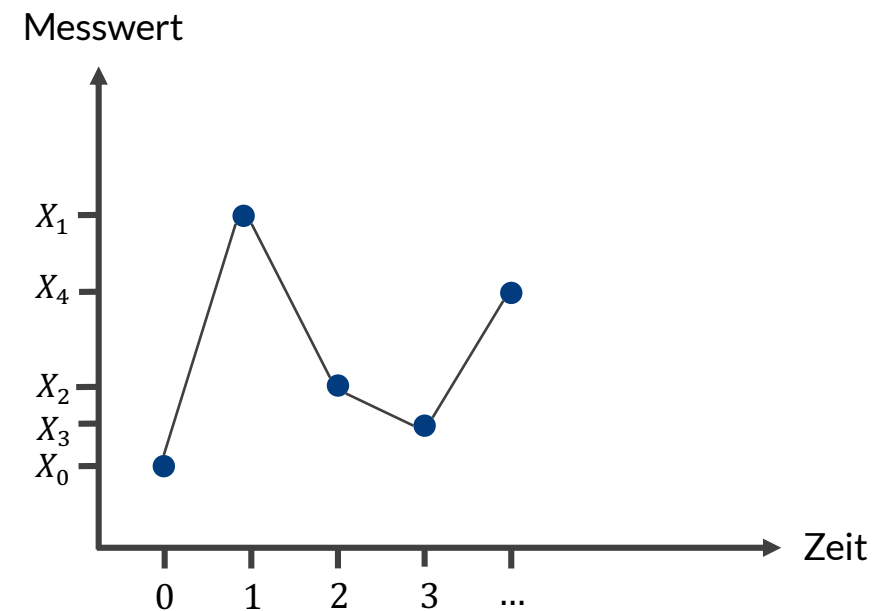
Ziel



Deep Dive

Was ist eine Zeitreihe?

1. Notiere den Messwert einer Beobachtung X
2. Notiere den Zeitpunkt $t = 0 \rightarrow X_0$.
3. Weitere Beobachtungen zu unterschiedlichen Zeiten mit **äquidistantem** Abstand $\rightarrow X_1, X_2, X_3, \dots$
4. Alle Beobachtungen zusammen ergeben die Zeitreihe $\rightarrow X = (X_0, X_1, X_2, \dots, X_t)$

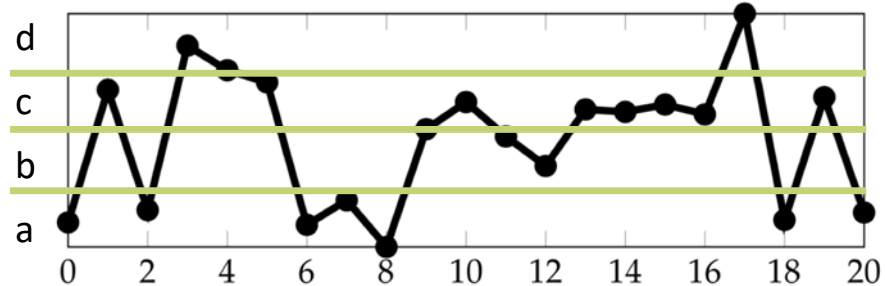


Abhängig von Zeit und Rauschen!

Symbolische Abstraktionen.

Symbolisierung

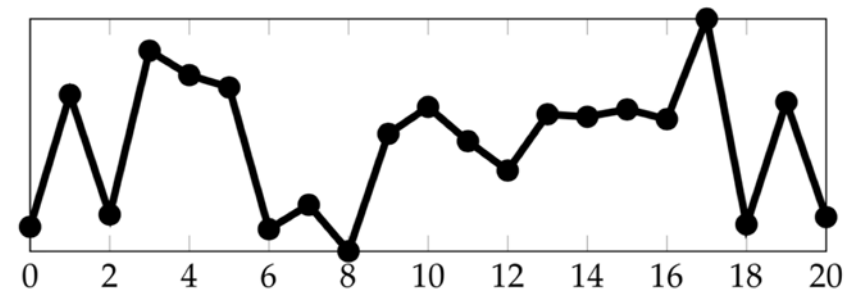
Klassisch:



Sequenz: acaddcaaaccbbcccdaca

- Unterteilung des Datenbereichs

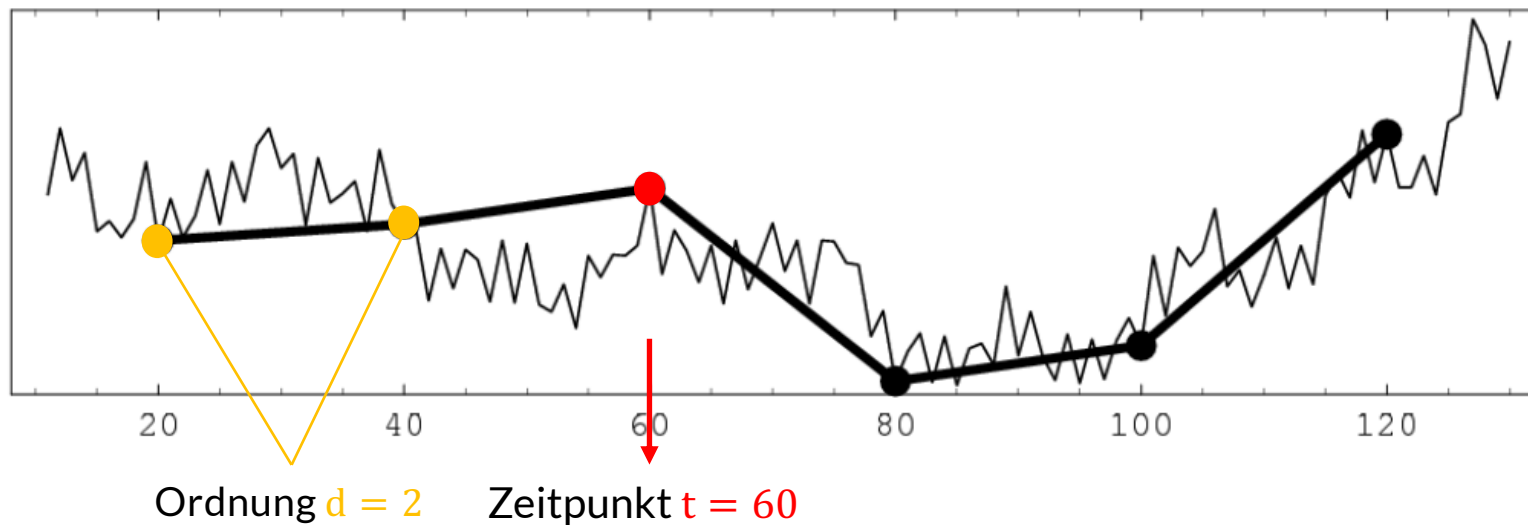
Ordinal:



Sequenz: ...

- Ordnungsrelationen

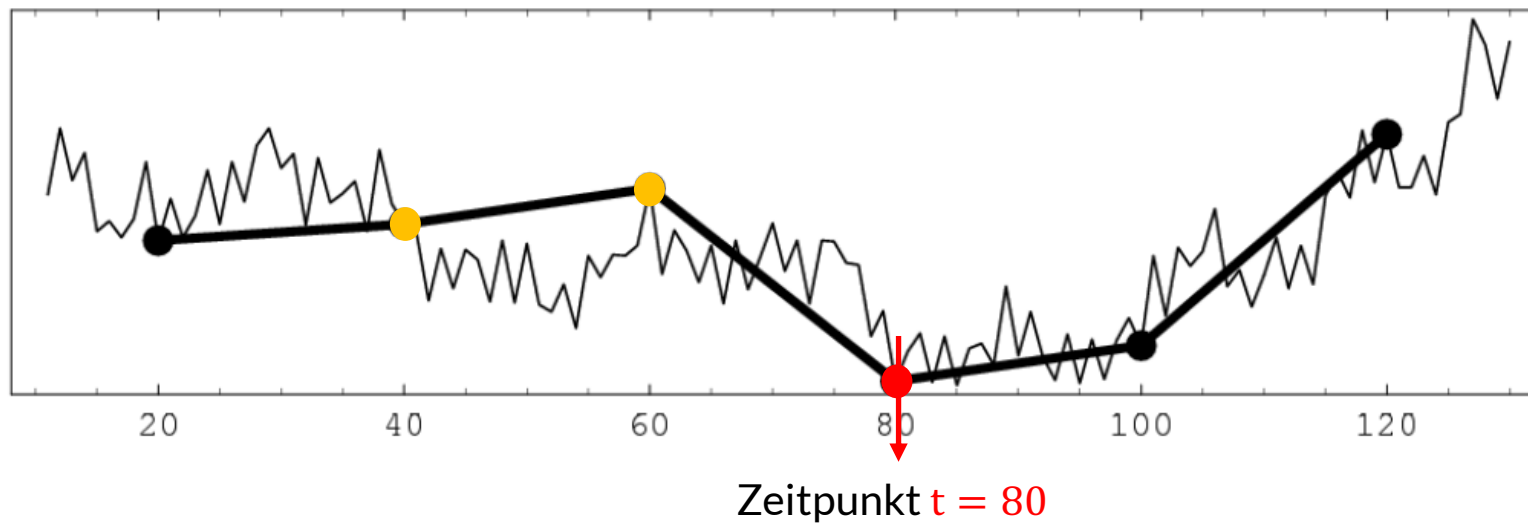
Ordinales Muster mit Delay $\tau = 20$ und Ordnung $d = 2$



$$x_{60-0 \cdot 20} > x_{60-1 \cdot 20} > x_{60-2 \cdot 20}$$

$$\pi_{2,20}(60) = (0, 1, 2)$$

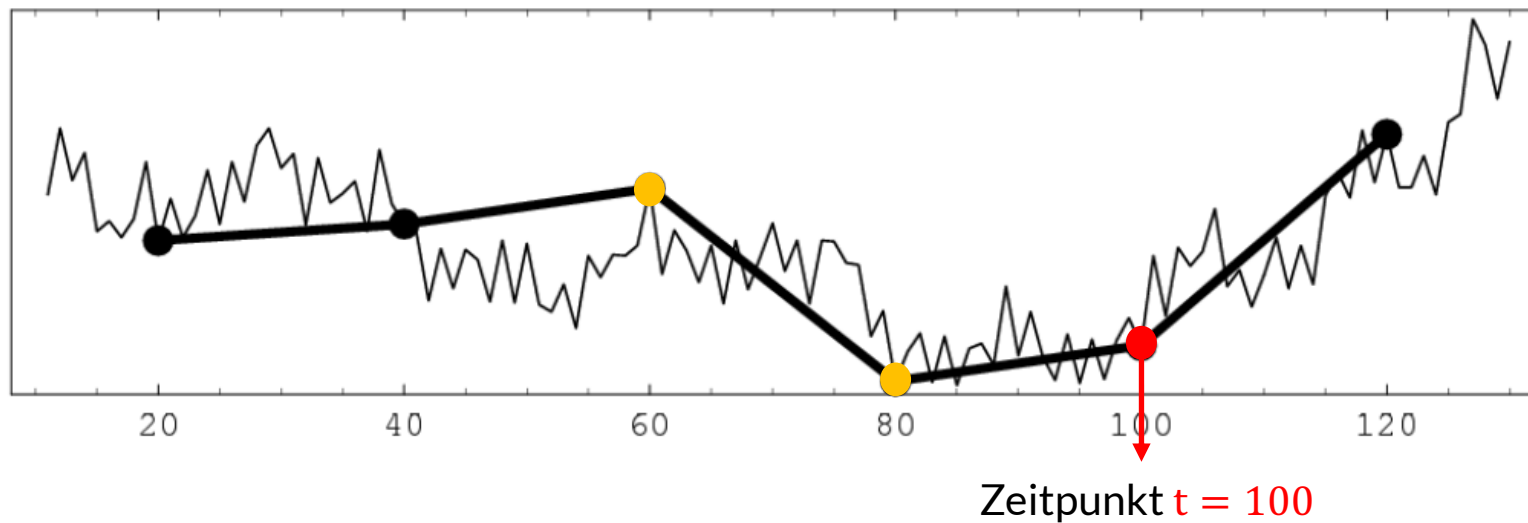
Ordinales Muster mit Delay $\tau = 20$ und Ordnung $d = 2$



$$x_{80-1 \cdot 20} > x_{80-2 \cdot 20} > x_{80-0 \cdot 20}$$

$$\pi_{2,20}(80) = (1, 2, 0)$$

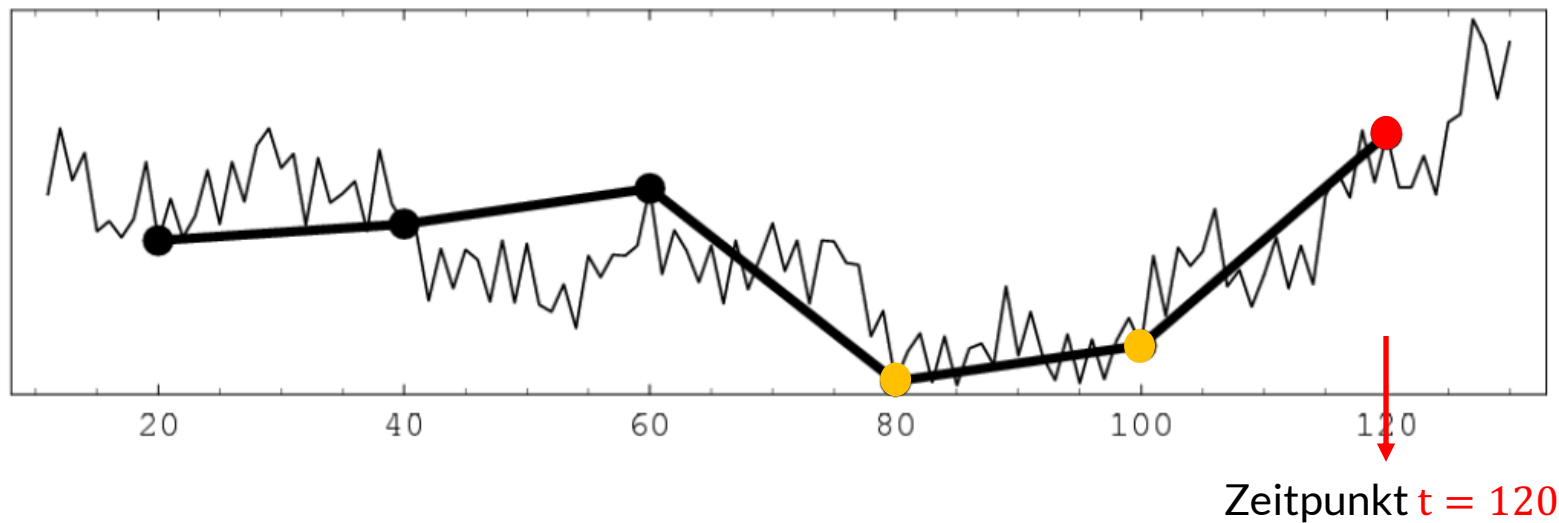
Ordinales Muster mit Delay $\tau = 20$ und Ordnung $d = 2$



$$x_{100-2 \cdot 20} > x_{100-0 \cdot 20} > x_{100-1 \cdot 20}$$

$$\pi_{2,20}(100) = (2, 0, 1)$$

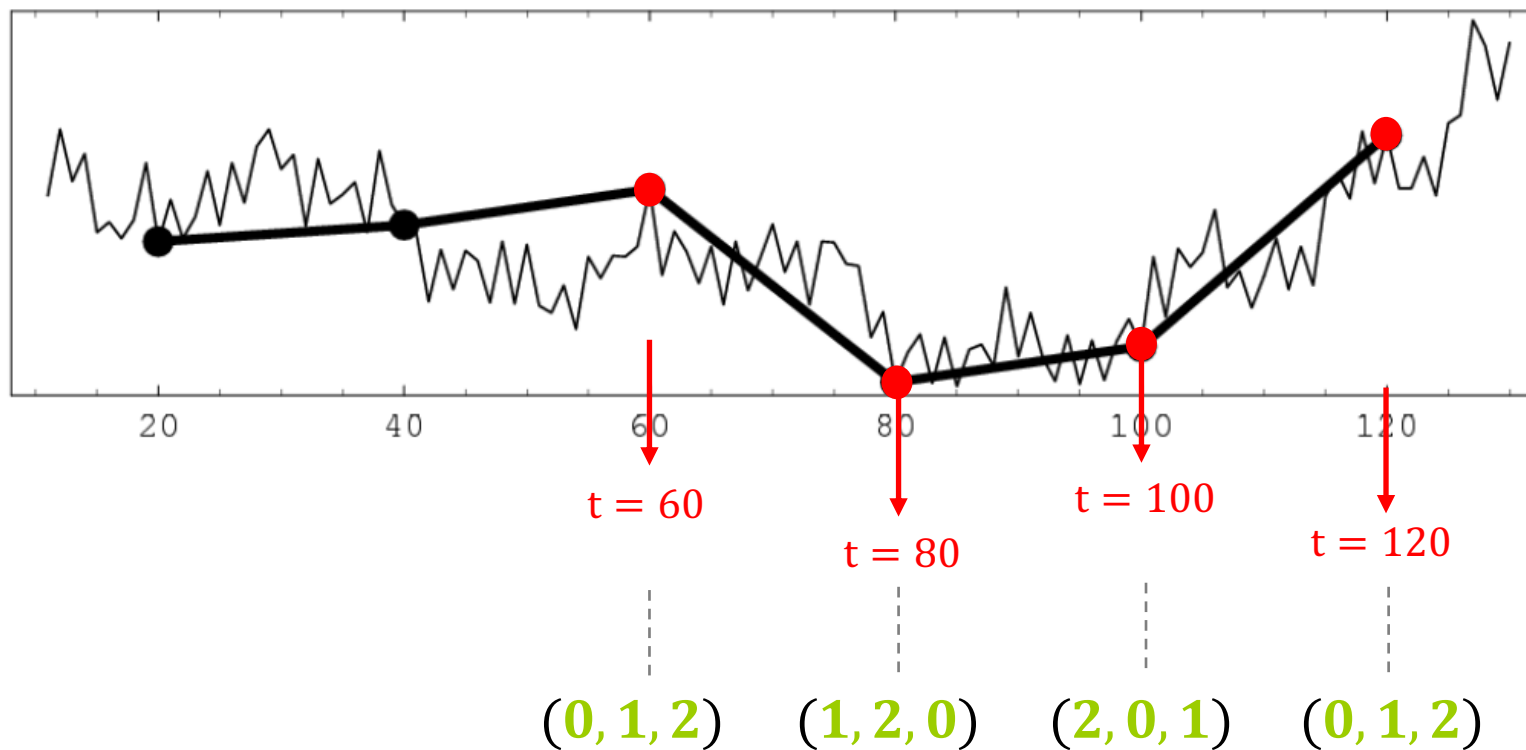
Ordinales Muster mit Delay $\tau = 20$ und Ordnung $d = 2$



$$x_{120-0 \cdot 20} > x_{120-1 \cdot 20} > x_{120-2 \cdot 20}$$

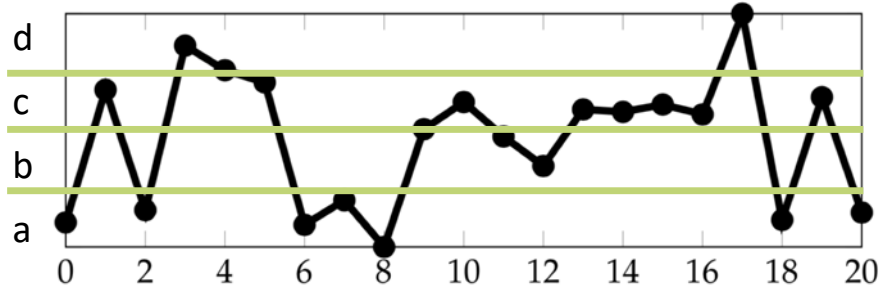
$$\pi_{2,20}(120) = (0, 1, 2)$$

Ordinales Muster mit Delay $\tau = 20$ und Ordnung $d = 2$



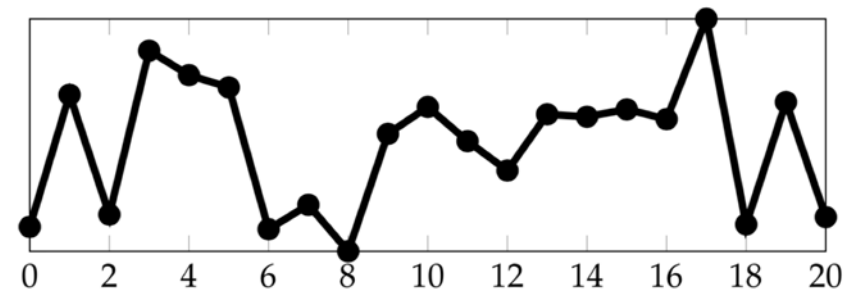
Symbolisierung von Zeitreihen

Klassisch:



Sequenz: acaddcaaaccbbcccdaca

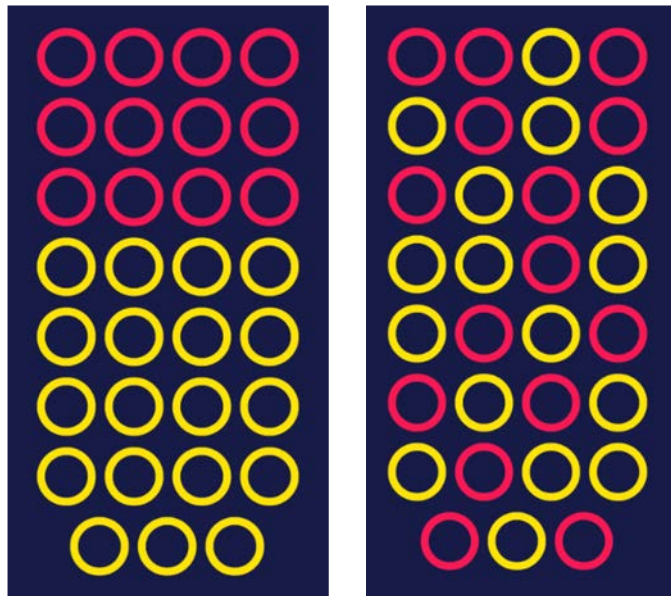
Ordinal:



Sequenzen: (2,0,1), (1,2,0), (2,0,1),....

Symbolisierungen von Zeitreihen, und nun?

Komplexität



acaddcaaaccbbccccdaca

abcdabcdabcdabcdabcd



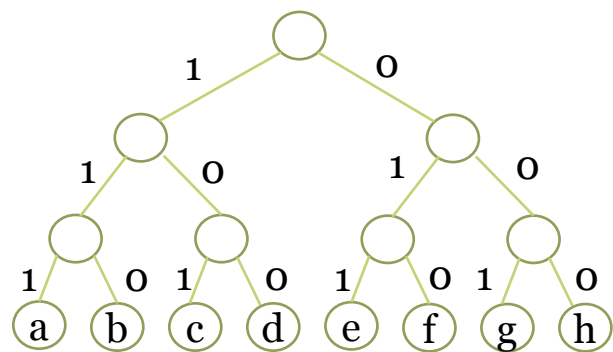
Klassischer Ansatz

Sequenz: acaddcaaaccbbcccdaca

Wie bestimmt man Komplexität?

Gleichverteilung

Benötigte Anzahl an Bits, um eine Symbol-Sequenz anzuzeigen oder zu übermitteln



Höhe
 $\log n$

$$p = \left(\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8} \right)$$

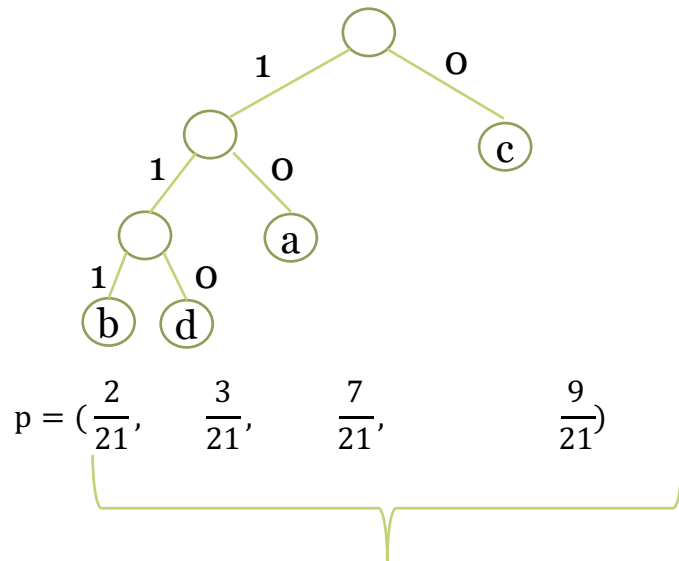
n Symbole, Gleichverteilung: $p_i = \frac{1}{n}$

$$\log n = \log \frac{1}{p} = \log 1 - \log p = -\log p$$

Wie bestimmt man Komplexität?

Keine Gleichverteilung

- › Nicht jedes Zeichen bzw. jede Nachricht ist gleichverteilt
- › Sequenz: **acaddcaaaccbbcccdaca**



n Zeichen, keine Gleichverteilung: $p_i \neq p_j$

Nachricht	Code	Länge	Erwartete Länge
b	111	3	0,286
d	110	3	0,429
a	10	2	0,667
c	0	1	0,429
Erwartete Länge			1,811

Entropie

Nichts Neues: Shannon (1948)

- (p_1, p_2, \dots, p_n) ist ein stochastischer Vektor, d.h.,
- $p_1, p_2, \dots, p_n \in [0,1]$ und
 - $\sum p_i = 1$



Gleichverteilung

$$\log n = -\log p$$

Keine Gleichverteilung

$$-\sum_{i=1}^n p_i \log p_i$$



Nutze das Auf und Ab!



Sequenzen: $(2,0,1)$, $(1,2,0)$, $(2,0,1)$,....

$$-\sum p_i \log p_i ,$$

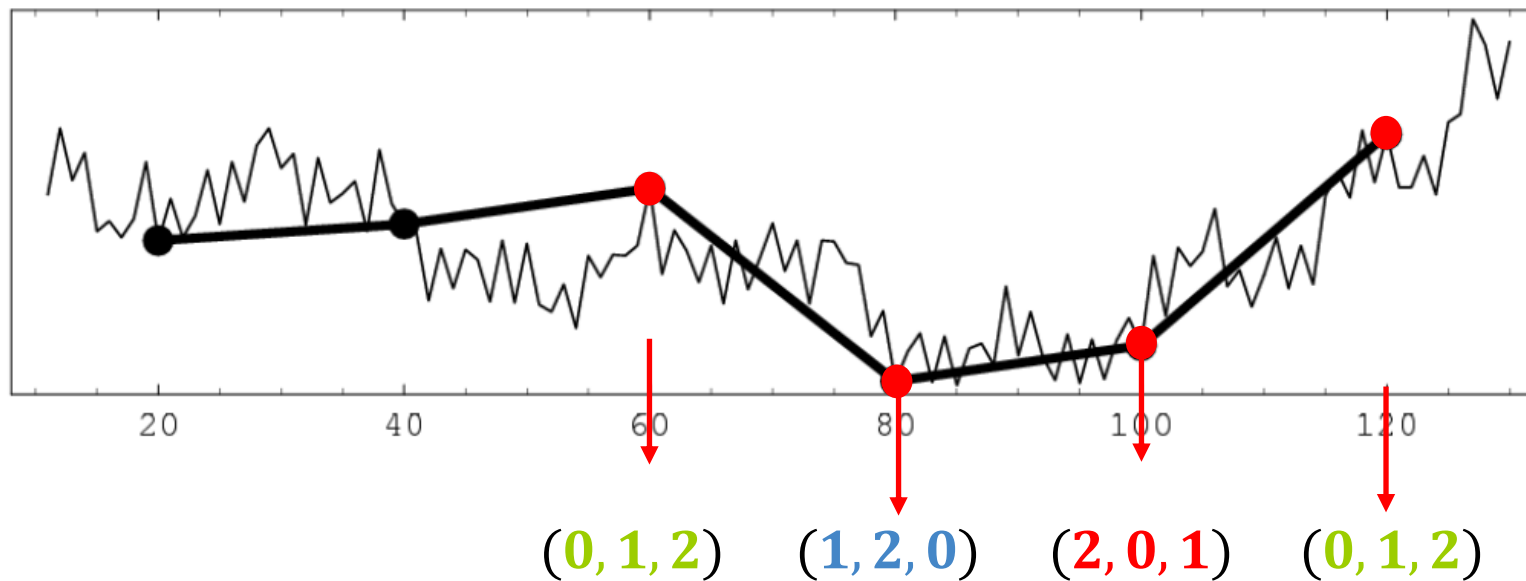
Permutationsentropie (PE)

= Shannon-Entropie ordinaler Muster

$$-\sum_{\pi \in \Pi} p(\pi) \log p(\pi),$$

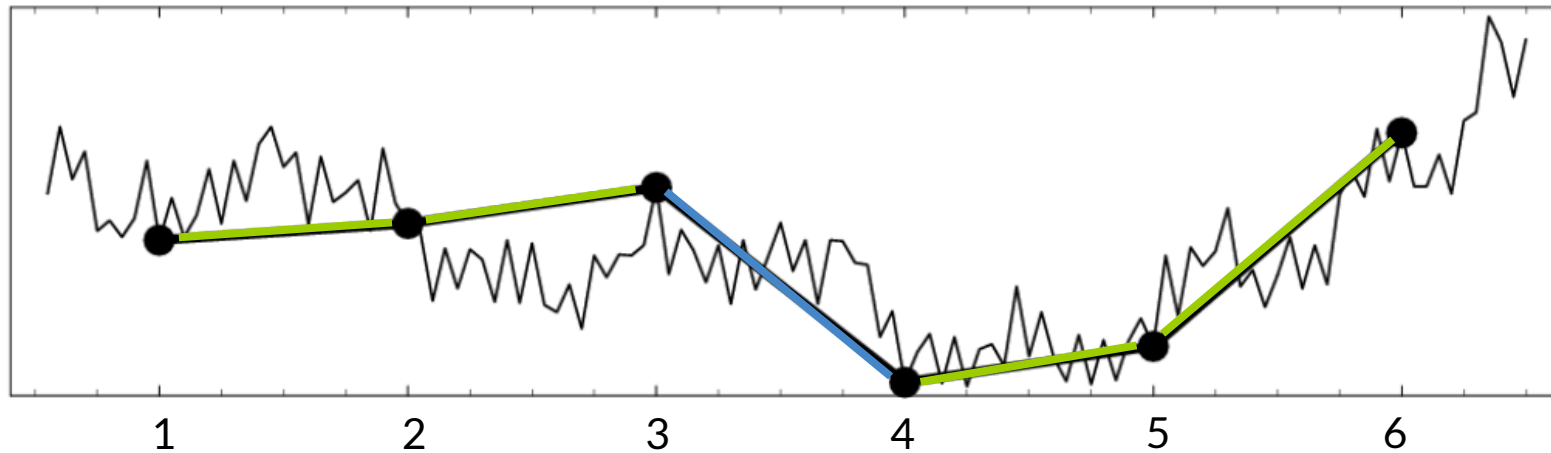
wobei $p(\pi)$ die relative Häufigkeit des ordinalen Musters π in der Zeitreihe ist.

Ordinales Muster mit Delay $\tau = 20$ und Ordnung $d = 2$



$$PE_2 = -\frac{2}{4}\log\left(\frac{2}{4}\right) - \frac{1}{4}\log\left(\frac{1}{4}\right) - \frac{1}{4}\log\left(\frac{1}{4}\right) = 1,5$$

Beispiel mit Ordnung $d = 1$



- 4 Fenster mit $x_t < x_{t+1}$, d.h., $\pi_1^{20} = (0,1)$
- 1 Fenster mit $x_t > x_{t+1}$, d.h., $\pi_1^{20} = (1,0)$

$$PE_1 = -\frac{4}{5} \log\left(\frac{4}{5}\right) - \frac{1}{5} \log\left(\frac{1}{5}\right) \approx 0,7219$$

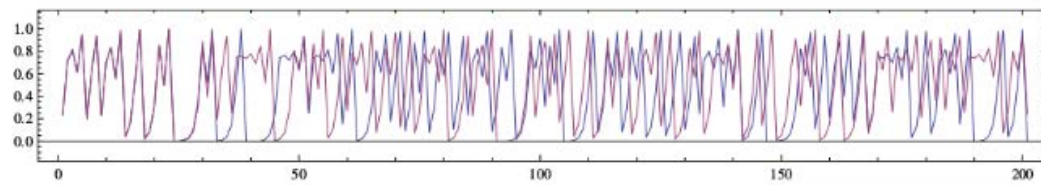
Fazit

Nur ein weiteres beschreibendes Feature?

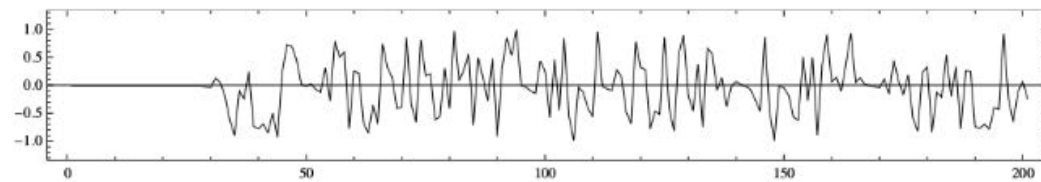
- „natürliches, intrinsisches“ Maß basierend auf **Informationen** statt Abständen
- Interpretierbares Auf und Ab des erzeugenden Systems
- Nachbarn werden berücksichtigt
- Konzeptionelles, interpretierbares Maß
- Kann um Bedingte W'keiten oder Amplituden erweitert werden
- Anwendung in Medizin (EEG, EKG, ...) oder Wirtschaft (Aktien, Populationen,...)

Chaos

- Ausgangssituation kann bei Wiederholung nicht identisch hergestellt werden
- Experiment landet nach kurzer Zeit in deutlich anderen Zuständen



two orbits of random points with distance 10^{-10}



distance of orbits

Wetter

Räuber-
Beute-
Modell

Tödliche
Embolien

Entstehung
böartiger
Tumore

Ausfall bestimmter
Hirnfunktionen
beim Schlaganfall

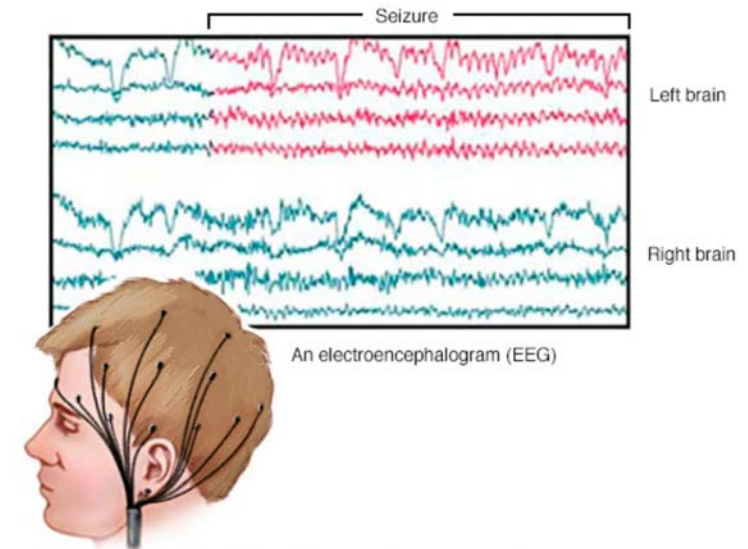
Anwendungen

Klassifikation + Detektion von epileptischen Anfällen

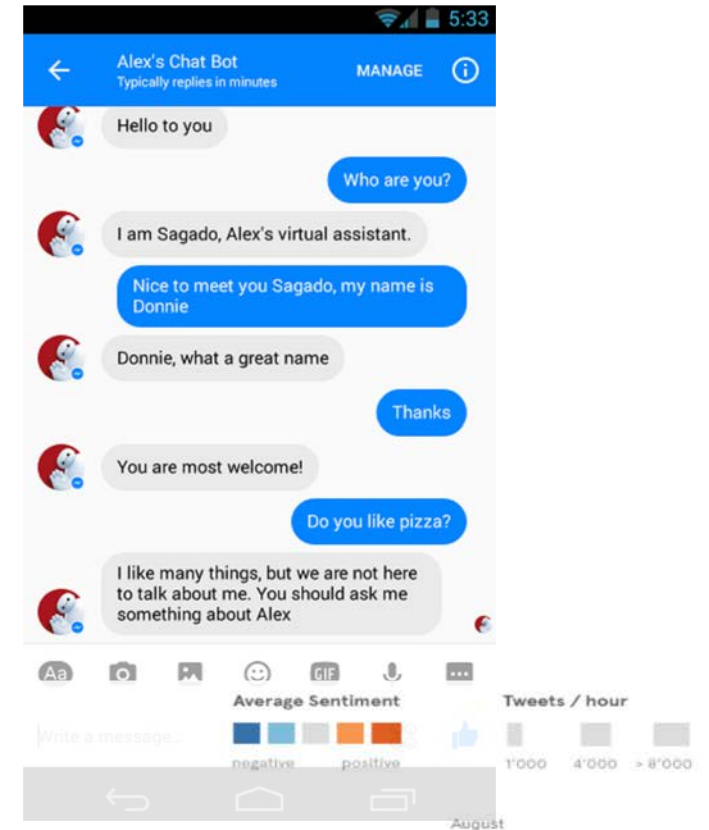
Daten:

- EEG-Datenbank
- Länge von 23.6s
- Sampling Rate 173.61 Hz.
- 5 Gruppen, je 100 Zeitreihen

	Gesund?	Wie gemessen?	Zustand
A	ja	Oberfläche	Offene Augen
B	ja	Oberfläche	Geschlossene Augen
C	nein	Intrakraniell, epileptogene Zone	Anfallsfreie Zeit
D	nein	Intrakraniell, gegenüberliegende Gehirnhälfte	Anfallsfreie Zeit
E	nein	intrakraniell	Anfallszeit

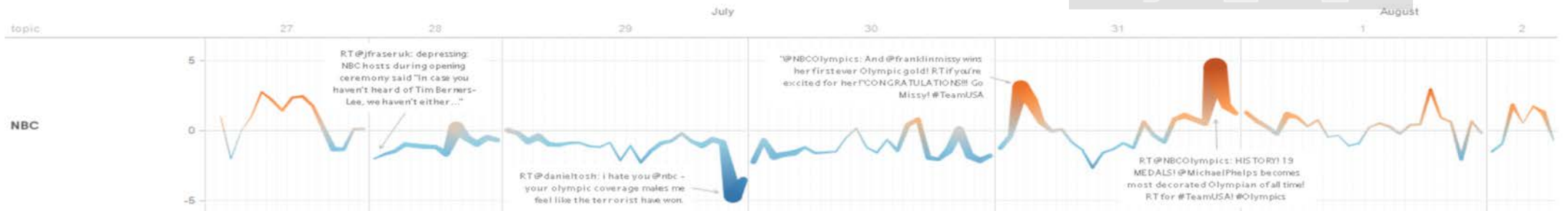


Wo spielt das "Auf und Ab" noch eine Rolle?



Tweet sentiments and volumes for selected topics

July 27, 08:00am - August 02, 12:00pm



<https://www.cmsattler.de/diktieren-spracherkennung-und-sprache-4-0/>, <http://truth-and-beauty.net/projects/emoto>, <https://medium.com/@5agado/building-a-personal-virtual-assistant-step-1-your-cv-as-a-chatbot-a4381fce6983>

Vielen Dank

Marisa Mohr

Machine Learning Engineer

inovex GmbH

Friesenweg 4 / Haus 13
22763 Hamburg

marisa.mohr@inovex.de

01523 3181 259

PANFABRIK



inovex