UNIVERSITÄT ZU LÜBECK
INSTITUT FÜR INFORMATIONSSYSTEME

Learning from Ups and Downs:

Multivariate Ordinal Pattern Representations for Time Series

Disputation

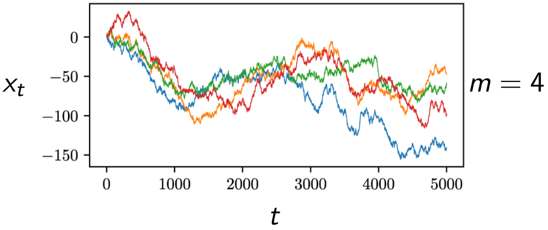
Marisa Mohr

Lübeck, 15. September 2022

Worum es heute geht: Zeitreihen.

- **Zeitreihen** sind Teil vieler realer Anwendungen, z. B.
  - Wetterdaten,
  - Aktienkurse,
  - Gehirn- und Herzaktivitäten

  – um nur ein paar zu nennen
- Zeitreihen sind (reellwertige) **Realisierungen** *dynamischer Systeme* oder *stochastischer Prozesse*



Es dreht sich alles um **Repräsentationen**!

- Die Vorhersagequalität eines datengesteuerten (KI- oder ML-)Modells ist in erster Linie durch die zum Training verwendeten Daten bestimmt

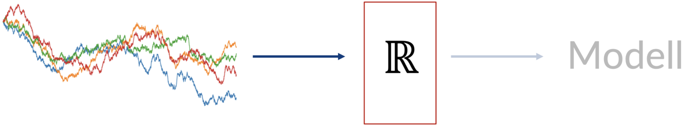
Aufgabe:

Dividiere 210 durch 6

Nun:

Dividiere *CCX* durch *VI*

- Eine *gute* Repräsentation ermöglicht oder vereinfacht eine nachfolgende Lernaufgabe

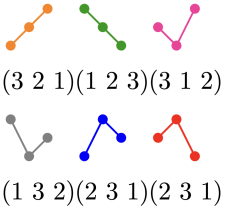


Ordinale Muster in Zeitreihen

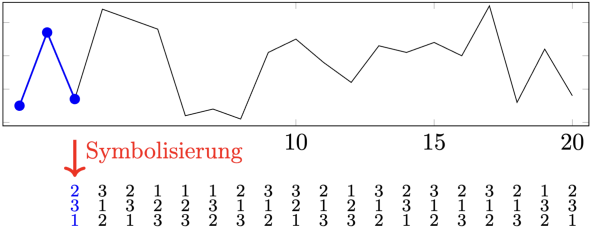
Bandt and Pompe (2002)

Definition

Ein Vektor $(x_1, \ldots, x_d) \in \mathbb{R}^d$ besitzt ein **ordinales Muster** $(r_1, \ldots, r_d) \in \mathbb{N}^d$ der Länge $d \in \mathbb{N}$ falls $x_{r_1} \geq \ldots \geq x_{r_d}$ bzw. $r_{l-1} > r_l$ im Fall $x_{r_{l-1}} = x_{r_l}$.



(a) Alle möglichen ordinalen Muster der Länge $d = 3$.



(b) Ordinale Musterbestimmung der Länge $d = 3$ zu allen Zeitpunkten $t \in [d, T]$.

Permutationsentropie
Bandt and Pompe (2002)

Definition

Die **Permutationsentropie (PE)** der Ordnung $d \in \mathbb{N}$ einer univariaten Zeitreihe $\mathrm{x} = (x_t)_{t=1}^{T}$, $T \in \mathbb{N}$ ist definiert durch

$$\mathrm{PE}_d(\mathrm{x}) = -\sum_{j=1}^{d!} p_j^d \cdot \log p_j^d, \tag{1}$$

wobei

$$p_j^d = \frac{\sum_{t \leq T}[(x_{t-(d-1)}, \ldots, x_t) \text{ hat Muster } j]}{T-(d-1)} \tag{2}$$

die Häufigkeit des ordinalen Musters j in der Zeitreihe ist, wobei $[x] = 1$, wenn x wahr ist, sonst 0.

Multivariate ordinale Musterrepräsentationen

In vielen Anwendungsbereichen werden multivariate Messungen durchgeführt:

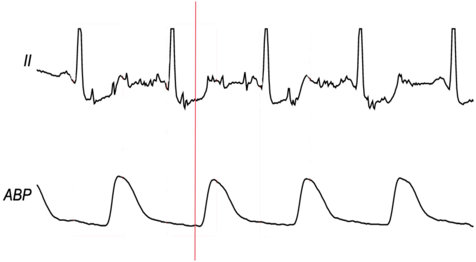


Abbildung 2: Medizinische multivariate Zeitreihen eines Patienten aus der MIMIC-III-Wellenformdatenbank mit der ID 3900006_0029, veröffentlicht von Johnson et al. (2016).

$$\begin{pmatrix} 6 \\ 1 \end{pmatrix} \overset{?}{\geq} \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

$$t \qquad\qquad t+1$$

Existierende und neue Strategien

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1T} \\ x_{21} & x_{22} & \cdots & x_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mT} \end{pmatrix} \to$$

PPE

MMSPE

MWPE

(a) Univariates ordinales Muster in der Zeit (Keller und Lauffer (2003), Morabito et al. (2012), Mohr et al. (2021)).

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1T} \\ x_{21} & x_{22} & \cdots & x_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mT} \end{pmatrix} \to \text{MvPE}$$

(b) Univariates ordinales Muster im Raum (He et al. (2016)).

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1T} \\ x_{21} & x_{22} & \cdots & x_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mT} \end{pmatrix}$$

$$\downarrow$$

$$\begin{pmatrix} x'_1 & x'_2 & \cdots & x'_T \end{pmatrix} \to$$

MPE-Eucl

MPE-Manh

MPE-Norm

MPE-PCA

(c) Univariates ordinales Muster über Dimensionsreduktion (Rayan et al. (2019), Mohr et al. (2020)).

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1T} \\ x_{21} & x_{22} & \cdots & x_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mT} \end{pmatrix} \to \text{MOPPE}$$

(d) Multivariates ordinales Muster (Mohr et al. (2020)).

Ups and Downs . . .

I . . . zur Schätzung von Modellparametern

- Marisa Mohr and Ralf Möller. **Ordering Principal Components of Multivariate Fractional Brownian Motion for Solving Inverse Problems.** In Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference 2021 (APSIPA-ASC), 2021
- Marisa Mohr, Florian Wilhelm, and Ralf Möller. **On the Behaviour of Weighted Permutation Entropy on Fractional Brownian Motion in the Univariate and Multivariate Setting.** The International FLAIRS Conference Proceedings, 34, 2021
- Marisa Mohr, Nils Finke, and Ralf Möller. **On the Behaviour of Permutation Entropy on Fractional Brownian Motion in a Multivariate Setting.** In Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference 2020 (APSIPA-ASC), 2020
- Marisa Mohr and Ralf Möller. **A Summary of Canonical Multivariate Permutation Entropies on Multivariate Fractional Brownian Motion.** Advances in Science, Technology and Engineering Systems Journal, 6(5), 2021

II . . . als Klassifikatoren im überwachten Lernen

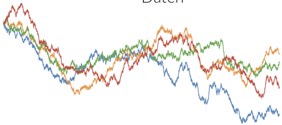
- Nils Finke, Marisa Mohr, Alexander Lontke, Marwin Züfle, Samuel Kounev, and Ralf Möller. **Recommendations for Data-Driven Degradation Estimation with Case Studies from Manufacturing and Dry-Bulk Shipping.** In Samira Cherfi, Anna Perini, and Selmin Nurcan (Eds.), Research Challenges in Information Science, 2021
- Marisa Mohr, Florian Wilhelm, Mattis Hartwig, Ralf Möller, and Karsten Keller. **New Approaches in Ordinal Pattern Representations for Multivariate Time Series.** In Proceedings of the 33rd International FLAIRS Conference (FLAIRS-33), 2020

III . . . zum Clustering im unüberwachten Lernen

- Nils Finke, Ralf Möller, and Marisa Mohr. **Multivariate Ordinal Patterns for Symmetry Approximation in Dynamic Probabilistic Relational Models.** In Guodong Long, Xinghuo Yu, and Sen Wang (Eds.), AI 2021: Advances in Artificial Intelligence, 2022
- Nils Finke and Marisa Mohr. **A Priori Approximation of Symmetries in Dynamic Probabilistic Relational Models.** In Stefan Edelkamp, Ralf Möller, and Elmar Rueckert (Eds.), KI 2021: Advances in Artificial Intelligence, 2021

I: Parameterschätzung
Inverses vs. Direktes Problem

beobachtete
Daten

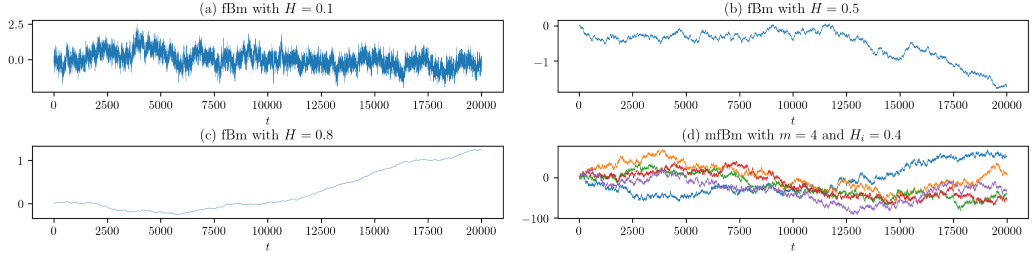
Modellparameter des erzeugenden Systems $\xleftarrow[\text{Problem}]{\text{inverses}}$  $\xrightarrow[\text{Problem}]{\text{direktes}}$ Wirkung und Prognosen

I: Parameterschätzung

Multivariate Fraktionale Brownsche Bewegung (mfBm)

- Klasse von stochastischen Prozessen, die
  - Gauß-verteilt und,
  - **selbstähnlich** (Hurst Parameter H) ist, und
  - stationäre Inkremente hat.
- Z. B. zur Modellierung von
  - Finanzdaten und Renditen,
  - Netzwerkverkehr,
  - u.v.m.

- $H < 1/2$: Negativ korrelierte Inkremente
- $H = 1/2$: gewöhnliche Brownsche Bewegung
- $H > 1/2$: Positiv korrelierte Inkremente
- $H \to 1$: Glattheit, weniger unregelmäßig und mehr trendig



I: Parameterschätzung
Bandt und Shiha (2007)

Univariates Beispiel $d = 3$:

$$p_{321} = \frac{1}{\pi} \cdot \arcsin 2^{H-1} =: u$$

Für alle weiteren ordinalen Muster gilt

$$p_j = \begin{cases} u & \text{falls } j = (123), (321), \\ (1-2u)/4 & \text{sonst} \end{cases}$$

I: Parameterschätzung
Mohr et al. (2020)

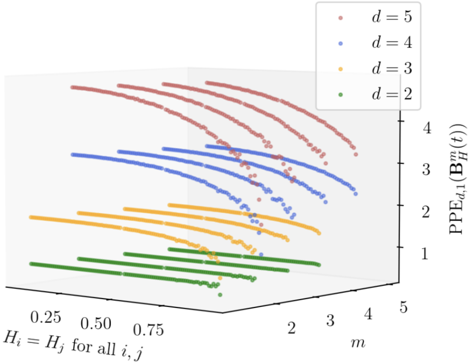
Multivariates Beispiel $d = 3, H_i = H_j$:

$$
\begin{aligned}
\mathrm{PPE}_3(\mathrm{B_H}(t)) \quad = \quad & -\Big(2 \cdot \frac{1}{\pi} \cdot \arcsin 2^{H_i - 1} \cdot \ln\left(\frac{1}{\pi} \cdot \arcsin 2^{H_i-1}\right) + \\
& \frac{1 - 2 \cdot \frac{1}{\pi} \cdot \arcsin 2^{H_i-1}}{4} \cdot \ln\left(\frac{1 - 2 \cdot \frac{1}{\pi} \cdot \arcsin 2^{H_i-1}}{4}\right)\Big)
\end{aligned} \tag{3}
$$

I: Parameterschätzung

$d = 3, 4, 5$	$H \in (0,1)^m$
PPE (Satz 6.2.2)	✓
MMSPE (Satz 7.2.2)	✓
MWPE (Satz 8.2.2)	✓
MPE-PCA (Kapitel 9.2)	✓
MOPPE (Lem. 10.1.2)	✓

Tabelle 1: Existenz eines Zusammenhangs zwischen MPE und *H*.



Zusammenfassung und Ausblick

I	› Eignung zur Schätzung von Parametern der mfBm	› Güte der Schätzung im Vergleich zu existierenden Ansätzen › Übertragbarkeit auf andere stochastische Prozesse
II	› Neue Repräsentationen schlagen existierende in verschiedenen Aufgaben zur Klassifikation	› Aufnahme in „Standard Data-Science-Werkzeugkästen"
III	› Algorithmus zum Clustering von Zeitreihen auf Basis von MOP (MOP_4SA)	› Anwendung im Kontext von Lifting

Die Suche nach der bestmöglichen Repräsentation bleibt ein spannendes Forschungsfeld.

Vielen Dank!

Marisa Mohr

www.marisa-mohr.de
mail@marisa-mohr.de
mohr@ifis.uni-luebeck.de

References

- Bandt und Pompe, Permutation Entropy: A Natural Complexity Measure for Time Series. Physical Review Letters, 88(17), 2002
- Amblard und Coeurjolly. Identification of the Multivariate Fractional Brownian Motion. IEEE Transactions on Signal Processing, 59(11), 2011
- Keller und Lauffer, Symbolic Analysis of High-Dimensional Time Series. International Journal of Bifurcation and Chaos, 13(09):2657–2668, 2003
- Johnson et al., MIMIC-III, a Freely Accessible Critical Care Database. Scientific Data, 3:160035, 2016
- Morabito et al., Multivariate Multi-Scale Permutation Entropy for Complexity Analysis of Alzheimer's Disease EEG. Entropy, 14(7), 2012
- He et al., Multivariate Permutation Entropy and its Application for Complexity Analysis of Chaotic Systems. Physica A: Statistical Mechanics and its Applications, 461, 2016
- Rayan et al., Multidimensional Permutation Entropy for Constrained Motif Discovery. In Ngoc Thanh Nguyen, Ford Lumban Gaol, Tzung-Pei Hong, and Bogdan Trawiński (Eds.), Intelligent Information and Database Systems, Lecture Notes in Computer Science, 2019