

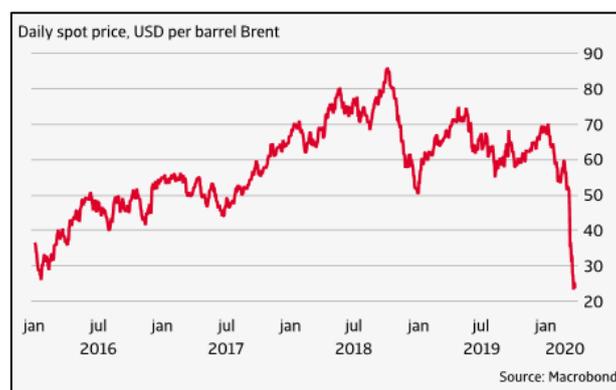
Prof. Dr.-Ing.habil.
Dirk Joachim Lehmann
Data Science in IoT
Fakultät für Informatik
di.lehmann@ostfalia.de
www.dirk-lehmann.de

Zeitreihenanalysetool

Hintergrund

Jede geordnete Folge von Datenwerten x_1, x_2, \dots, x_m entspricht einer konkreten *Reihe* von Daten, man spricht auch von ordinalen Daten. Werden diese ordinalen Daten entlang konkreter Zeitpunkte t_1, t_2, \dots, t_m betrachtet, bilden sich ordinalen Zeit-Daten-Paare $(x_1, t_1), (x_2, t_2), \dots, (x_m, t_m)$. Wir sprechen in diesem Fall von einer (univariaten) *Zeitreihe*.

Zeitreihen können sich diskret oder kontinuierlich verhalten: Eine diskrete Zeitreihe kennt zwischen zwei aufeinander folgenden konkreten Zeitpunkten t_m und t_l keine weiteren. Die Zeitpunkte sind somit abzählbar. Ein Beispiel sind die Spielzustände bei einem Schachspiel. Im ersten Zug – also zum ersten Zeitpunkt t_1 – hat das Spiel eine bestimmte Startkonfiguration. Im zweiten Zug – also zum zweiten Zeitpunkt t_2 – wurde eine Spielfigur gesetzt (z.B. das Pferd) und das Spiel hat eine neue Konfiguration. Zwischen zwei Zügen/Zeitpunkten gibt es keine weiteren \Rightarrow diskrete Zeitreihe. Eine kontinuierliche Zeitreihe besteht aus überabzählbaren Zeitpunkten. Eine physikalische Messung über die Zeit ist ein konkretes Beispiel für eine kontinuierliche Zeitreihe, z.B. die Messung des Luftdrucks an einem bestimmten Punkt im Tagesverlauf. Aus praktischen Gründen sind kontinuierliche Messungen als diskrete Punkte in der Praxis gegeben. Dies kann zu Verwirrungen mit diskreten Zeitreihen führen. Bitte machen Sie sich in einem solchen Fall klar, dass es sich dann um diskrete Samples einer kontinuierlichen Domäne handelt.

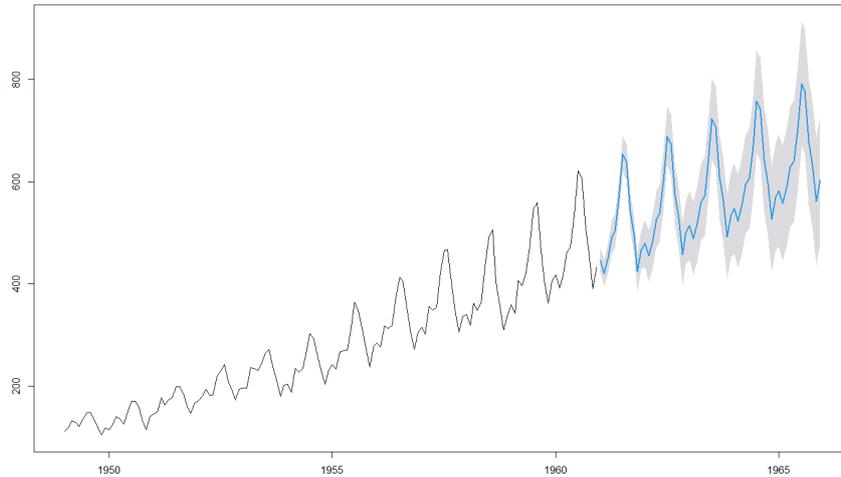


Zeitreihe: Zeitlicher Verlauf einer Messgröße

Zeitreihen heben sich gegenüber anderen Daten dahingegen ab, das ihnen – der Theorie nach - eine inhärente Kausalität innewohnt, also eine Ursache-Wirkungs-Relation: Der Idee nach ergeben sich „spätere Ereignisse/Daten“ kausal ursächlich aus den Eigenschaften „früherer Ereignisse/Daten“.

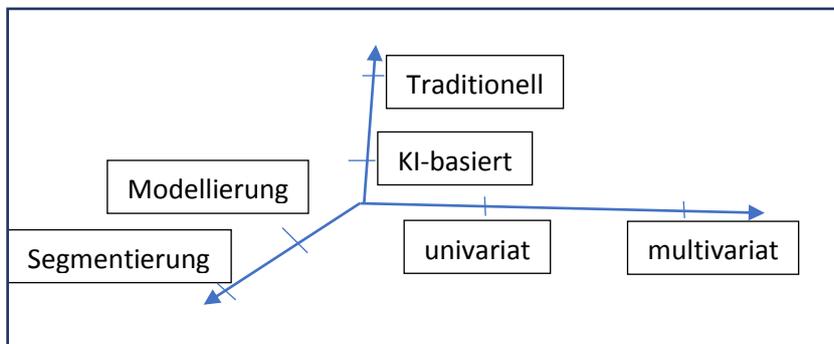
Zeitreihen, welche uns im Alltag begegnen, sind Börsendaten, Elektroenzephalogramm-Daten (EEG), Daten aus Motorsteuergeräten, Wetter- und Klimadaten u.v.m.

Die Analyse von Zeitreihen ist Gegenstand von Studien und Forschung. Ziel ist es (gedämpfte, verstärkende, konstante) zyklische Muster zu identifizieren, den zugrundeliegenden Datengenerierungsprozess zu verstehen, oder zukünftige Werte mit begrenztem Unsicherheitsbalken zu präzidieren.



Blaue Kurve modelliert zukünftige Werte der schwarzen Zeitreihe (Unsicherheit der Prädiktion – grauer Bereich – nimmt stetig zu)

Wir können zwischen traditionellen – und KI-basierten Analysemethoden unterscheiden, sowie zwischen der Analyse durch Modellierung der Zeitreihe und Analyse durch Segmentierung der Zeitreihe. Zudem können zeitreihen mit einem Merkmal auftreten (univariat) oder mit mehreren (multivariat). Folgende Abbildung fasst dies zusammen:



Hinzu kommen noch weitere Ansätze, wie vergleichende Metriken zw. Zeitreihen und Ähnliches. Im Folgenden werden einige Methoden kurz andiskutiert.

Traditionelle Univariate Zeitreihenanalyse über Modellierung

In der modellierenden Analyse wird die Zeitreihe durch mathematische Modelle nachgebildet. Hierbei wird oft ein stochastischer Prozess in dem Zeitreihenerzeugungsprozess unterstellt [1]. Diesbezüglich werden Komponenten angenommen aus denen sich die Zeitreihe zusammensetzt, wie Signalanteile, Rauschanteile, oder skalenbasierte Anteile, welche sich zur Zeitreihe zusammensetzen.

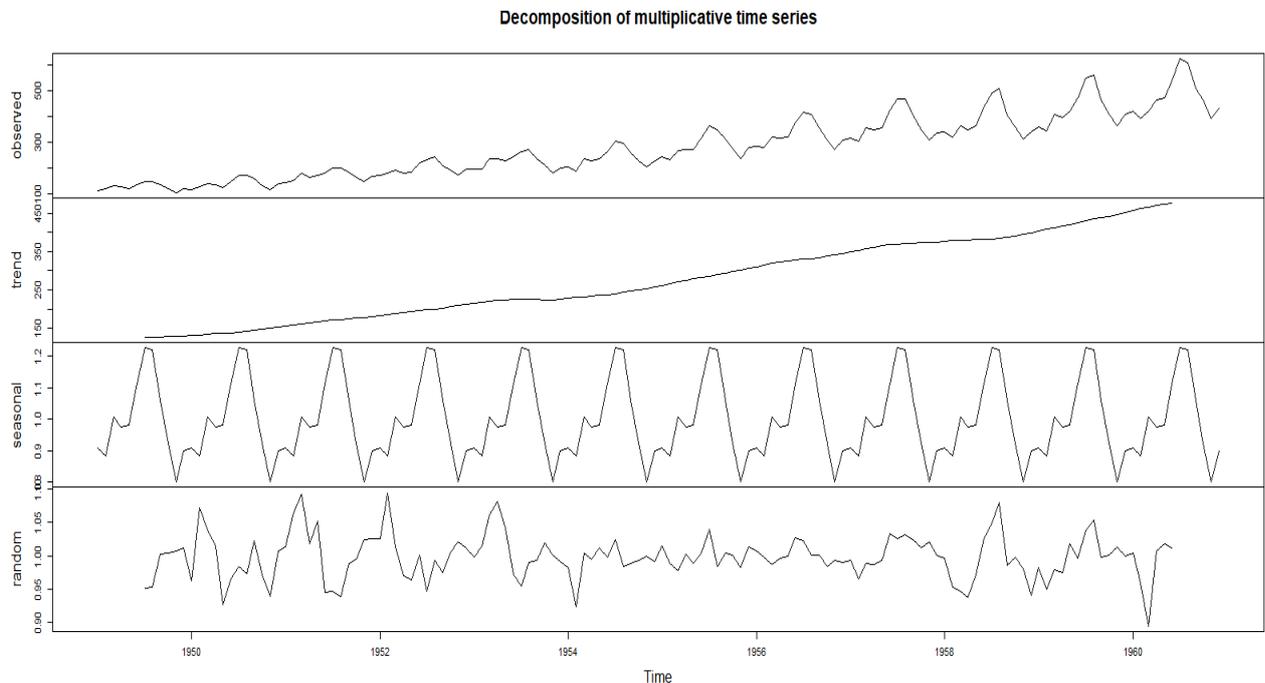
ARMA-Modell (AutoRegressive-Moving Average)

$$y_t = c + \epsilon_t + \sum_{j=1}^q b_j \epsilon_{t-j}$$

Die Zeitreihe y_t wird modelliert durch die konstant c , das aktuelle gleitende Mittel e_t , und der durch b_j gewichteten gleitenden Mittel der früheren q -Perioden e_{t-j} . Erweitert werden diese Modelle durch Autoregressives (AR-Modelle). Weitere Varianten sind ARIMA, ARMAX, VARMA.

Saisonale ARMA-Modelle

Saisonale ARMA-Modelle zerlegen eine Zeitreihe (hier observed) in Teil-Zeitreihen unterschiedlicher Perioden aus welchen es sich zusammensetzt (hier trend, seasonal, random) so das gilt: $\text{observed} = \text{trend} + \text{seasonal} + \text{random}$



Polynomiale (und nicht lineare Regressionsmodelle) Regressionsmodelle

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_h X^h + \varepsilon$$

Die Zeitreihe Y wird über ein Polynom h -ten Grades modelliert, durch eine Daten-basierte geeignete Wahl der Parameter in B_i .

Traditionelle Univariate Zeitreihenanalyse über Ordinale Analysen (Ordinale Zeitreihenanalyse)

In der ordinalen Zeitreihenanalyse werden die Ordnungsrelationen zwischen den zeitlichen Datenwerten beschrieben; insofern also inwieweit die Werte steigen/fallen oder sich konstant verhalten. Dadurch ergeben sich ordinale Muster (welche beispielsweise nun über Strings beschrieben werden können), welche sich einfach zueinander und untereinander vergleichen lassen und metrisch und statistisch weiter charakterisiert werden können (z.B. über die Permutationsentropie).

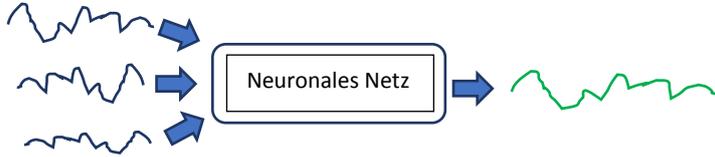
KI-basierte Univariate Zeitreihenanalyse

Bei der univariaten Zeitreihenanalyse werden Neuronale Netze genutzt um die Zeitreihe (blau) prädiktiv fortzusetzen (orange). Die Herausforderung besteht darin eine entsprechende Topologie des Neuronalen Netzes zu finden, welche korrekt prädiziert.



KI-basierte Multivariate Zeitreihenanalyse

Bei der multivariaten Zeitreihenanalyse werden unterschiedliche Eingangs-Zeitreihen genutzt um eine (oder mehrere) weitere (vorliegende) Zeitreihe in ihrem Datengenerierungsprozess zu modellieren.



Traditionelle Univariate Zeitreihenanalyse über Segmentierung

Ein alternative Analysetechnik für Zeitreihen ist es die Reihe in charakteristische Segmente zu überführen. Dadurch können unterschiedliche Zeitreihen untereinander effektiv auf periodische Ähnlichkeiten hin analysiert werden, zudem ist es auch möglich selbstähnliche Abschnitte einer Zeitreihe auf sich selbst zu untersuchen.



Weitere häufige Analysemethoden sind Ansätze über Autokorrelationen, Kreuzkorrelationen, Frequenz-basierte Ansätze (Fourier etc.) Registrierungen von Zeitreihen aufeinander u.v.m.

Projektziel

Ziel in diesem Projekt ist es ein Software-Werkzeug zu entwickeln, das methodisch auf dem aktuellen Stand der Forschung ist, um Zeitreihenanalysen durchführen zu können. Dies umfasst ebenfalls die Entwicklung geeigneter Nutzer-Interfaces und Visualisierungen mit geringer Nutzungsbarriere.

Referenzen

- [1] Box, G.E.P; Jenkins, G.M.: *Time Series Analysis - Forecasting and Control*, San Francisco: Holden Day, 1970
- [2] <https://de.wikipedia.org/wiki/Zeitreihenanalyse>
- [3] <https://statistik.econ.kit.edu/download/Deskriptive12.pdf>
- [4] https://epub.ub.uni-greifswald.de/frontdoor/deliver/index/docId/1907/file/diss_wittfeld_katharina.pdf,
- [5] https://dewiki.de/Lexikon/Zeitreihenanalyse#Ordinale_Zeitreihenanalyse

Angebot

Im Rahmen ihres *Praxisprojektes*, *Masterseminars*, *Masterprojektes*, ihrer *Bachelorarbeit*, *Masterarbeit* oder ähnlichen Studienleistungen - wie z. B. einem *interdisziplinären Digitalisierungsprojekt* - können sie gerne diese Aufgabe bearbeiten

Melden Sie sich gerne bei mir unter: di.lehmenn@ostfalia.de

Aufgaben/Arbeitspakete

Das Projekt besteht aus klar voneinander abgrenzbaren Arbeitspaketen (AP):

- **AP1) Dashboard / User-Interface**

- o [Eingabe](#): Datenadressen für Zeitreihen
- o [Ausgabe](#): Visualisierung der Zeitreihen, Analyseinformationen, und Widgets für entsprechende Aktionen

Beispiele für in diesem Paket umzusetzende Features:

- Darstellen und Visualisierungen von Zeitreihen
- Editieren von Parametrisierungen
- Visualisieren von Modellen, Segmente etc. von Zeitreihen
- Darstellung von Metainformationen, wie Änderungsverläufe für eingegebenen Analyseergebnisse oder Nutzerinformationen etc.
- Darstellen von Widgets für Actions etc.
- ...

- **AP2) Traditionelle Methoden zum Modellieren univariater Zeitreihen**

- o [Eingabe](#): Datenstruktur für eine konkrete Periode einer univariaten Zeitreihe, Modell- und Systemparameter
- o [Ausgabe](#): Traditionelle Modelle welche die Zeitreihe modellieren

Beispiele für in diesem Paket umzusetzende Features:

- Implementieren von ARMA für univariate Zeitreihen
- Implementieren von Saisonales-ARMA für univariate Zeitreihen
- Implementieren von ARIMA für univariate Zeitreihen
- Implementieren von ARMAX für univariate Zeitreihen
- Implementieren von Polynomieller Regression für univariate Zeitreihen
- Implementieren von weiteren nicht-linearen Regressionen univariater Zeitreihen
- Implementation von Multiskalen-Modellen univariater Zeitreihen
- ...

- **AP3) Traditionelle Methoden zur ordinalen Analyse univariater Zeitreihen**

- o [Eingabe](#): Datenstruktur für eine konkrete Periode einer univariaten Zeitreihe, Modell- und Systemparameter
- o [Ausgabe](#): Ordinale Modelle der Zeitreihe

Beispiele für in diesem Paket umzusetzende Features:

- Implementation von Transformationen in ordinale Muster von Zeitreihen
- Implementieren von Permutationsentropie
- ...

- **AP4) KI-Methoden zum Modellieren univariater Zeitreihen**

- o [Eingabe](#): Datenstruktur für eine konkrete Periode einer univariaten Zeitreihe, Modell- und Systemparameter
- o [Ausgabe](#): KI-Modelle welche univariate Zeitreihen modellieren

Beispiele für in diesem Paket umzusetzende Features:

- Implementation von Rekurrente Neuronale Netze für univariate Zeitreihen
- Implementation von LSTM für univariate Zeitreihen
- Implementation weitere Topologien für Neuronale Netze für univariate Zeitreihen
- ...

- **AP5) KI-Methoden zum Modellieren multivariater Zeitreihen**

- o [Eingabe](#): Datenstruktur für eine konkrete Periode einer multivariater Zeitreihe, Modell- und Systemparameter
- o [Ausgabe](#): KI-Modelle welche multivariate Zeitreihen modellieren

Beispiele für in diesem Paket umzusetzende Features:

- Implementation von Rekurrente Neuronale Netze für multivariate Zeitreihen
- Implementation von LSTM für univariate multivariate Zeitreihen
- Implementation weitere Topologien für Neuronale Netze für univariate Zeitreihen
- ...

- **AP6) Maße für die Analyse univariater/multivariater Zeitreihen**

- o [Eingabe](#): Datenstruktur für eine konkrete Periode einer univariaten Zeitreihe, Modell- und Systemparameter
- o [Ausgabe](#): KI-Modelle welche multivariate Zeitreihen modellieren

Beispiele für in diesem Paket umzusetzende Features:

- Implementation von Autokorrelationsfunktionen univariater Zeitreihen
- Implementation von Kreuzkorrelationen multivariater Zeitreihen

- ...
- **AP7) Maße für die Analyse univariater/multivariater Zeitreihen**
 - [Eingabe](#): Datenstruktur für eine konkrete Periode einer univariaten Zeitreihe, Modell- und Systemparameter
 - [Ausgabe](#): Angewandte Maße über univariate/multivariate Zeitreihen

Beispiele für in diesem Paket umzusetzende Features:

 - Implementation von Autokorrelationsfunktionen univariater Zeitreihen
 - Implementation von Kreuzkorrelationen multivariater Zeitreihen
 - ...
- **AP8) Frequenzbasierte Analyse univariater/multivariater Zeitreihen**
 - [Eingabe](#): Datenstruktur für eine konkrete Periode einer univariaten/multivariater Zeitreihe, Modell- und Systemparameter
 - [Ausgabe](#): Angewandte Frequenzanalysen über univariater/multivariater Zeitreihen

Beispiele für in diesem Paket umzusetzende Features:

 - Implementation von Fourier-basierten Analysen für Zeitreihen
 - ...
- **AP9) Segmentierungsansätze univariater Zeitreihen**
 - [Eingabe](#): Datenstruktur für eine konkrete Periode einer univariaten Zeitreihe, Modell- und Systemparameter
 - [Ausgabe](#): Beschreibungen charakteristischer Segmente von Zeitreihen, Vergleichsmaße zw. charakteristischer Segmente zw. unterschiedlicher Zeitreihen

Beispiele für in diesem Paket umzusetzende Features:

 - Segmentierungen von Zeitreihen
 - Clusterisierung von Zeitreihen
 - Vergleiche von Segmenten zw. unterschiedlichen Zeitreihen
 - ...
- **AP10) Modelgüte-Validierung für Modelle von Zeitreihen**
 - [Eingabe](#): Modelle von Zeitreihen, Evaluierungsdaten, Modell- und Systemparameter
 - [Ausgabe](#): Beurteilung von Modelgüte, Prädiktionsgüten, Anpassungs- und Evaluierungsgüten betrachteter Modelle/Segmente/Maße über Zeitreihen

Beispiele für in diesem Paket umzusetzende Features:

 - Implementation von Evaluierungskonzepte für Zeitreihenmodelle
 - Implementation von Validierungskonzepte von Zeitreihenmodelle
 - Implementation von Prädiktionsgütenabschätzung von Zeitreihenmodellen
 - Implementation von Schätzungen zu Überanpassungen/Unteranpassungen
 - ...
- **AP11) Daten-Accessor**
 - [Eingabe](#): Datenpfade, Datenstream-Quellen-Adressen,
 - [Ausgabe](#): Speicher- und Lade-Operationen

Beispiele für in diesem Paket umzusetzende Features:

 - Persistente Speicherung von Datenanalyseergebnissen
 - Laden/Speichern von Datensätze von denen Zeitreihenanalysen durchgeführt werden sollen
 - Streamen von Datensätze von denen Zeitreihenanalysen durchgeführt werden sollen
 - Zeitreihen-Datenstrukturen für die Arbeit mit geladenen Zeitreihen
 - Sampling und Cropping Schemes für Zeitreihen-Datenstrukturen
 - ...

Hinweise: Nicht die gesamten Daten sollen geladen werden und modelliert werden in den APs 2-10, sondern nur solche Perioden und Samples, welche durch den Nutzer ausgewählt wurden. Wie diese Ausgewählt werden, wir in AP 11 implementiert.

Zu den Arbeitspaketen hinzu kommen notwendige Recherchetätigkeiten, Make-Or-Buy-Entscheidungen, Beachtung von Lizenzfragen, Aspekte der Continuous Integration und des Code-Managements, Dokumentationsaufgaben, Fragen zum Aufsetzen/Deployment und der Migration von Entwicklersystemen (und zum Projektmanagement), wie sie in der Softwareentwicklung üblich sind.

Im Rahmen einer Projektbearbeitung wird nicht erwartet, dass unmittelbar alle APs bearbeitet werden können. Je nach Umfang Ihrer zu erbringenden Studierendenleistung können Teilaspekte einzelner

APs bearbeitet werden. Den konkreten, jeweiligen Umfang stimmen wir im Vorfeld gerne gemeinsam ab.

Vorkenntnisse

Es ist hilfreich – aber keine Voraussetzung – wenn Sie Vorkenntnisse/Interesse mitbringen in

- Softwareentwicklung
 - Python
 - Datenanalyse mit Zeitreihen
 - Data Engineering
-

Vorarbeiten

Kurzeinführung in Python:

https://www.youtube.com/watch?v=x_kYpwi1L1k

OnboardingProzess

Alle wichtigen Zugänge einrichten um die Arbeit am Projekt aufnehmen zu können:

<http://46.38.235.241/webpage/dirkfiles/misc/onboarding/OnboardingProzess.pdf>

Organisatorisches

Projektmanagement per Scrum in Trello:

<https://trello.com/b/qfC1TK5b/zeitreihenanalysetool>

Projektcode-Management per GitHub:

<https://github.com/DirkJLehmann/Zeitanalysetool.git>

Projektdokumentations-Management:

https://docs.google.com/document/d/17wcGUT_nVP1sZcatY1adP8Wgff6dNbM760d1ct5Sdd8/edit?usp=sharing

Bei Interesse melden Sie sich bitte unter:

Prof. Dr.-Ing.habil.
Dirk Joachim Lehmann
Data Science in IoT
Fakultät für Informatik
di.lehmann@ostfalia.de