

# Cleaning von Zeitreihen

Von der Anomalieerkennung zur Anomalienreparatur

---

Jose Rodriguez Parra Flores  
Klaus-Johan Ziegert

16. September 2019



# Gliederung

---

- 1 Einführung
- 2 Grundlagen
- 3 Iterative Minimum Repairing
- 4 Evaluierung
- 5 Schluss

# Einführung

---

- 1 Einführung
  - Motivation
  - Zielsetzung

2 Grundlagen

3 Iterative Minimum Repairing

4 Evaluierung

5 Schluss

# Motivation

## Messgeräte liefern unzuverlässige Daten

- GPS Tracker sind nahe von Gebäuden unzuverlässig
- Sensoren sind empfindlich gegenüber äußere Einflüsse
  - Z.B. starker Fall der Temperaturen bei einem Windzug



Abbildung: GPS-Tracking auf dem Campus der Tsinghua Universität [1]

# Motivation

## Umgang von unzuverlässigen Daten mit Anomalieerkennung

### ① Unzuverlässige Datenpunkte entfernen

- Ausreißer werden entfernt 😊
- Entfernen aufeinanderfolgende Fehler machen Ergebnis unbrauchbar **oder** werden als solche ggf. nicht entfernt 😞

### ② Unzuverlässige Datenpunkte reparieren

- Einzelne Ausreißer werden leicht korrigiert 😐
- Aufeinanderfolgende Fehler werden zu stark verändert (In der Praxis liegen die Messungen nahe bei den korrekten Werten) 😐

# Motivation

## Hinzunahme von korrekt markierten Werten

- ① Markierung durch den Benutzer
  - Z.B. markiert der Benutzer in beliebigen Zeitabständen seinen aktuellen Standort
- ② Präzise Messgeräte liefern in längeren Zeitabstände korrekte Werte



# Zielsetzung

## Ziel der Arbeit

- ① Berücksichtigung der markierten Werte in der Anomalieerkennung
  - Aufeinanderfolgende Fehler sollen besser abgeschätzt werden
- ② Anomalienreparatur mit den Minimum-Change-Prinzip vereinbaren
  - Keine drastische Veränderungen der Messwerte
- ③ Neue Anomalienreparatur hinsichtlich Berechnungslaufzeit, Ergebnisgenauigkeit usw. optimieren
- ④ Neue Anomalienreparatur mit unterschiedlichen Einstellungen mit den anderen Verfahren empirisch vergleichen

# Grundlagen

---

- 1 Einführung
- 2 Grundlagen**
  - Problemstellung
  - Anomalien
  - Reparatur durch Anomalieerkennung
- 3 Iterative Minimum Repairing
- 4 Evaluierung
- 5 Schluss



# Problemstellung

## Zeitreihenreparatur

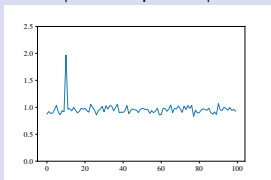
- Gegeben:
  - Eine fehlerbehaftete Messung  $x = x[1], \dots, x[n]$
  - Eine unvollständige, aber dafür ausschließlich korrekte Messung  $x^{\text{truth}}$
- Gesucht:
  - Reparatur  $y$  mit minimalen RMS-Fehler  $\Delta(x^{\text{truth}*}, y)$
  - $\Delta(x^{\text{truth}*}, y) = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i^{\text{truth}*} - y_i)^2}$

# Anomalien

## Anomalien

- Wikipedia: Abweichung von der Regel
- Werte  $x_i$  mit Abweichung  $\tau$  (Bsp.  $\tau = 2\sigma$ ):

$$|x_i - x_i^{\text{truth}^*}| > \tau$$



# Reparatur durch Anomalieerkennung

## Autoregressive Modell $AR(p)$

- Lineare Regression der letzte  $p$  Werte:

$$x'_t = \sum_{i=1}^p \phi_i x_{t-i} + \epsilon_t$$

- Reparatur:

$$y_t = \begin{cases} x'_t & \text{falls kein Label und } |x'_t - x_t| > \tau \\ x_t & \text{sonst} \end{cases}$$

# Reparatur durch Anomalieerkennung

## Autoregressives exogenes Modell $ARX(p)$

- Exogenes Variabel  $y$

$$y'_t = x_t + \sum_{i=1}^p \phi_i(y_{t-i} - x_{t-i}) + \epsilon_t$$

- $y'_t$  Mögliche Reparatur:

$$y_t = \begin{cases} y'_t & \text{falls kein Label und } |y'_t - x_t| > \tau \\ y_t & \text{sonst} \end{cases}$$

- Fakten

# Iterative Minimum Repairing

- 1 Einführung
- 2 Grundlagen
- 3 Iterative Minimum Repairing**
  - allgemeines IMR
  - Matrix-Pruning IMR
  - Incremental-Computation IMR
- 4 Evaluierung
- 5 Schluss

# allgemeines IMR

## Blank

- Fakten
- Fakten
- Fakten

# Matrix-Pruning IMR

## Blank

- Fakten
- Fakten
- Fakten

# Incremental-Computation IMR

## Blank

- Fakten
- Fakten
- Fakten



# Evaluierung

---

1 Einführung

2 Grundlagen

3 Iterative Minimum Repairing

**4 Evaluierung**

- Ordnung
- Schwellenwert
- maximale Anzahl von Iterationen
- Markierungsrate

5 Schluss

# Ordnung

## Blank

- Fakten
- Fakten
- Fakten

# Schwellenwert

## Blank

- Fakten
- Fakten
- Fakten

# maximale Anzahl von Iterationen

## Blank

- Fakten
- Fakten
- Fakten

# Markierungsrate

## Blank

- Fakten
- Fakten
- Fakten

# Schluss

---

- 1 Einführung
- 2 Grundlagen
- 3 Iterative Minimum Repairing
- 4 Evaluierung
- 5 Schluss**
  - Zusammenfassung und Ausblick
  - Literatur

# Zusammenfassung und Ausblick

## Zusammenfassung

- Was wurde getan?

# Zusammenfassung und Ausblick

## Zusammenfassung

- Was wurde getan?


## Ausblick

- Wie könnten zukünftige Arbeiten aussehen?



# Literatur I

---

 Shaoxu song - tsinghua university.

 Aoqian Zhang, Shaoxu Song, Jianmin Wang, and Philip S Yu.

Time series data cleaning: From anomaly detection to anomaly repairing.

*Proceedings of the VLDB Endowment*, 10(10):1046–1057, 2017.