

Cleaning von Zeitreihen

Von der Anomalieerkennung zur Anomalienreparatur

Jose Rodriguez Parra Flores
Klaus-Johan Ziegert

13. 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 unzuverlässig in der Nähe von Gebäuden
- 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

Ansätze mit den Umgang von unzuverlässigen Daten

- ① Unzuverlässige Datenpunkte mit der Anomalieerkennung entfernen
 - Einzelne Ausreißer werden entfernt 😐
 - Entfernen langer Folgen von Ausreißer machen Ergebnis unbrauchbar 😞
 - Lange Folgen von Ausreißer werden als solche ggf. nicht erkannt 😞
- ② Unzuverlässige Datenpunkte mit der Anomalieerkennung reparieren
 - Einzelne Ausreißer werden korrigiert 😊
 - Lange Folgen von Ausreißer werden zu stark verändert (widerspricht den Minimum-Change-Prinzip) 😐
 - Lange Folgen von Ausreißer werden als solche ggf. nicht verändert 😞

Motivation

Hinzunahme von als wahr markierte Werte

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



Zielsetzung

Ziel der Arbeit

- Einfließen der markierten Werte in die Anomalieerkennung
 - lange Folgen von Ausreißer sollen identifiziert werden
- Anomalieerkennung mit den Minimum-Change-Prinzip vereinbaren
 - keine drastische Veränderungen der Messwerte
- Einfließen der markierten Werte in die anderen State-of-Art Verfahren
- Neues Verfahren mit den anderen Verfahren empirisch vergleichen

Grundlagen

- 1 Einführung
- 2 Grundlagen**
 - Problemstellung
 - bekannte Verfahren
- 3 Iterative Minimum Repairing
- 4 Evaluierung
- 5 Schluss

Problemstellung

Blank

- Fakten
- Fakten
- Fakten [2]

bekannte Verfahren

Blank

- Fakten
- Fakten
- 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.