Bachelorarbeit Informatik

Stefan Bechert

4. März 2016

Inhaltsverzeichnis

1 Einleitung **3**
- 1.1 Motivation 3
- 1.2 Zielsetzung 3
- 1.3 Gliederung 3

2 Grundlagen **3**
- 2.1 Plotly 3
- 2.2 Zeitreihen, Repräsentationen und Operationen 3
  - 2.2.1 Definition Array 3
  - 2.2.2 Definition Zeitreihe 4
  - 2.2.3 Repräsentationen 4
  - 2.2.4 Operationen - allgemein 5
- 2.3 Pivot Tabellen 7

3 Anforderungsanalyse und Konzept **7**

4 Umsetzung **7**

5 Zusammenfassung und Evaluation **7**
- 5.1 Auswertung 7
- 5.2 Ausblick 7

1 Einleitung

1.1 Motivation

Zeitreihen,

Wofuer braucht man diese Arbeit? Beispiele aus Industrie.

1. Einsatz in vielen industriellen Projekten.
2. Fehleranalyse.
3. Vergleiche in flexiblen zeitlichen Dimensionen.

1.2 Zielsetzung

Welche Ziele sollen erreicht werden?

1.3 Gliederung

Wie ist die Arbeit aufgebaut? Warum ist sie das?

2 Grundlagen

Welche theoretischen und praktischen Grundlagen sind zur Umsetzung dieser Arbeit erforderlich?

2.1 Plotly

Wieso Plotly? Grundlagen?

2.2 Zeitreihen, Repräsentationen und Operationen

2.2.1 Definition Array

Sei $n \in \mathbb{N}$. Ein **Array** $arr = [a_0, a_1, ..., a_{n-1}]$ ist eine endliche Abfolge von Werten, die ueber ihren Index i referenziert werden. Es gilt $arr[i] = a_i$, wobei $i \in \{0, 1, ..., n-1\}$.

Die **Länge** eines Arrays $|arr|$ entspricht der Anzahl der Werte in arr, also $|arr| = n$.

Sei arr ein Array. Falls ein $i \in \{0, 1, ..., |arr| - 1\}$ existiert sodass $arr[i]$ selbst ein Array ist so heisst arr **geschachtelt** und $arr[i]$ Unterarray von arr. Ansonsten heisst arr **nicht geschachtelt**.

Sei arr ein Array. Die Menge $\boldsymbol{\chi(arr)}$ ist die Menge aller Arrays in arr und deren rekursiv erreichbaren Unterarrays, die selbst nicht geschachtelt sind. (Eigentlich muesste man das besser ausdruecken)

2.2.2 Definition Zeitreihe

Eine **äquidistante Zeitreihe** ist ein Tripel

$$D^2 = (V, i, s) \tag{1}$$

wobei

$$V = [v_0, v_1, ..., v_{n-1}] \tag{2}$$

ein Array aus n Kennwerten darstellt. i beschreibt den zeitlichen Abstand zwischen zwei Kennwerten in Millisekunden und s repräsentiert den Starzeitpunkt der Zeitreihe in Millisekunden seit dem 01.01.1970.

2.2.3 Repräsentationen

Die Definition in 2.2.2 gilt im folgenden als die initiale Repräsentation einer Zeitreihe in der Anwendung. (Passt hier eig nicht, sollte eher zu Implementierung)

2D - Repräsentation Die Repräsentation nach 2.2.2 ist eine 2-dimensionale Darstellung der Kennzahlen, da nur ein Index jede Kennzahl eindeutig referenziert.

3D - Repräsentation Jede Zeitreihe kann auch in Abhängigkeit von zwei unterschiedlichen Zeitdimensionen angegeben werden.

Sei $D^2 = (V, i, s)$ eine Zeitreihe in 2D-Repräsentation. So lässt sich diese Zeitreihe auch darstellen in 3D-Repräsentation als

$$D^3 = (V', i, s) \tag{3}$$

mit

$$V' = [[v_0, v_1, ..., v_{t_1-1}]_0, [v_0, v_1, ..., v_{t_2-1}]_1, ..., [v_0, v_1, ..., v_{t_m-1}]_m] \tag{4}$$

wobei

$$\sum_{k=1}^{m} t_k = |V|. \tag{5}$$

Falls

$$t_1 = t_2 = ... = t_m \tag{6}$$

gilt, so nennt man V' eine **homogene**, ansonsten eine **inhomegene** 3D-Repräsentation von D.

Folgerung Homogen 3D Sei D^3 als homogen gegeben. So folgt:

$$\sum_{k=1}^{m} t_k = t * m = |V|. \tag{7}$$

nD - Repräsentation, Ueberlegung Allgemein kann man jede Zeitreihe in n vielen Dimensionen darstellen (da beliebig kleine Zeitschritte), allerdings erreicht man dadurch irgendwann keine neuen Informationen mehr sondern ausschliesslich Kennwertduplizierung.

2.2.4 Operationen - allgemein

Sei K^n die Menge aller Zeitreihen in nD-Repraesentation.

Unterteilung $\forall n \geq 2$ gilt:
Sei $D^n = (V, i, s)$ eine Zeitreihe in nD-Repraesentation.
Eine Unterteilung δ_n ist eine Abbildung

$$\delta_n : K^n \to K^{n+1}. \tag{8}$$

Sei $D^n = (V, i, s)$ homogen und $\delta(D^n) = D^{n+1} = (V', i, s)$. Eine Unterteilung heisst **homogen** falls ein $n \in \mathbb{N}$ exisitiert, sodass

$$\forall arr \in \chi(V) : |arr| = n. \tag{9}$$

Aggregationssregeln Sei $D^n = (V, i, s)$ eine Zeitreihe in nD-Repraesentation.

1. Eine Zeitreihe ist genau dann aggregierbar, wenn $n \geq 3$.
2. Sei $D^n = (V, i, s)$ homogen. Fuer
$$\phi(D^n) = D^{n-1} = (V', i', s) \tag{10}$$
gilt
$$i' = |MIA(D^n)| * i \tag{11}$$
.
3. Sei $D^n = (V, i, s)$ inhomogen. Fuer
$$\phi(D^n) = D^{n-1} = (V', i', s) \tag{12}$$
gilt
$$i' = [l_0, l_1, ..., l_{|MIA|-1}] \tag{13}$$
wobei $i'[0] = l_0 = |MIA[0]|$.

Aggregation $\forall n \geq 3$ gilt:
Sei $D^n = (V, i, s)$ eine Zeitreihe in nD-Repraesentation.

Eine Aggregation ϕ_n ist eine Abbildung
$$\phi_n : K^n \to K^{n-1}. \tag{14}$$

Unterteilung	**Aggregation**
D^2	D^2
$\downarrow$	$\uparrow$
D^3	D^3
$\downarrow$	$\uparrow$
...	...
$\downarrow$	$\uparrow$
D^{n-1}	D^{n-1}
$\downarrow$	$\uparrow$
D^n	D^n

2.3 Pivot Tabellen

Sei $D = (V, i, s)$ eine Zeitreihe. So kann D auch als Tabelle dargestellt werden:

Zeit in msec	s	$s + i$	$s + 2 * i$	...	$s + (n - 1) * i$
Kennwert	v_0	v_1	v_2	...	v_{n-1}

3 Anforderungsanalyse und Konzept

Was soll die Software am Ende koennen? Wie sollte man rangehen?

4 Umsetzung

Tatsaechliche Umsetzung: Probleme, Entscheidungen,

Wichtige Forderungen TODO aber merrken Dass die Teilarrays in der 3D Repraesentation die gleich Groesse haben ist fuer den spaeteren Render-Prozess von hoher Wichtigkeit, da andernfalls Daten in der Darstellung verloren gehen. Nichts destotrotz kann eine Zeitreihe auch in ungleiche Teile zerteilt werden (z.B. Monate), die allerdings zur Darstellung auf jeden Fall aggregiert werden muessen um die Forderung einzuhalten.

5 Zusammenfassung und Evaluation

5.1 Auswertung

Wurden die Ziele erfuellt? Wenn ja zeigen. Wenn nein, wieso nicht? Zeitpunkt: Nach der Fertigstellung der software.

5.2 Ausblick

Was koennte man aufbauend auf dieser Arbeit noch machen?