

Übung Algorithm Engineering

Toni Barth und Max Haarbach

29. Juni 2018

1. Heuristiken

1.1. Heuristik 1: Zufallsdrehungen

Die erste Heuristik führt eine bestimmte Anzahl an Drehungen, die von der Grösse der Instanz abhängt, an zufällig ausgewählte Knoten aus. Dieser Vorgang wird wiederum je nach Grösse der Instanz mehrfach durchgeführt und am Ende die Sortierung mit dem geringsten Abstand als Ergebnis ausgegeben.

1.2. Heuristik 2: Optimal Leaf Ordering

Die zweite Heuristik nutzt das Verfahren, das Bar-Joseph und weitere für eine möglichst schnelle und optimale Sortierung von hierarchisch geclusterten Datensätzen entwickelt haben [BGJ01]. Dabei wird folgender rekursiver Ansatz verfolgt: Sollen Kosten für einen bestimmten Knoten berechnet werden, setzen sich diese aus den Kosten der beiden Kindknoten und dem Abstand der beiden inneren Blätter dieser beiden Teilbäume. Sofern der Knoten, für den Kosten berechnet werden sollen, ein Blatt bzw. einen Datensatz darstellt, betragen dessen Kosten 0. Dies ist daher das Rekursionsende. Begonnen wird üblicherweise mit dem Wurzelknoten, da man dadurch am Ende auch die gesamten Abstandskosten berechnet hat.

2. Ziele

Durch die Experimente sollen sowohl

- die Laufzeiten der Heuristiken bei unterschiedlichen Grössenordnungen bezüglich der Anzahl der Testobjekte als auch
- die Güte aufgrund der Ähnlichkeiten zu den jeweiligen Originalbildern

ermittelt und verglichen werden.

3. Faktoren

Beim „Leaf-ordering“ sind lediglich 2 Faktoren von Bedeutung:

Zum Einen bestimmt die GrösSe der Bilder, die im Endeffekt die Anzahl der Testobjekte widerspiegelt, die Laufzeit der Heuristiken. Zum Anderen spielt auch deren Struktur oder Art eine Rolle, die sich allerdings schwer in konkrete MessgrösSen oder Werte fassen lassen.

4. Testinstanzen

Gemäss der **Faktoren** werden auch die Testinstanzen, die durch Grauwert-Bilder realisiert sind, in die entsprechenden Kategorien unterteilt:

- GrösSen:
 - 10
 - 50
 - 100
 - 500
 - 1000
 - 2500
 - 5000
- Arten:
 - (symmetrische) Testbilder
 - Fotos der realen Welt
 - Farb- bzw. Grau-Übergänge

5. Ergebnis-Qualität

Als MaSS für die Qualität der Ergebnisse wird die Summe der Abstände aller benachbarten Blatt-paare genutzt, sodass bei 10 Blättern 9 Abstände zu addieren sind. Die Abstände wiederum werden durch den euklidischen Abstand der entsprechenden Spaltenvektoren berechnet. Die Messwerte der originale Testinstanzen werden später eingefügt, da die Implementierung dazu noch fehlt. Für die gemischten und sortierten Instanzen sind die Werte im **Anhang A** zu finden.

Aufgelistet werden erstmal nur Werte für die Instanzen der GrösSen 10, 50 und 100, da ab 500 die Laufzeit der zweiten Heuristik stark zunimmt. Es wurden insgesamt 3 Durchläufe durchgeführt, in denen jeweils eine Testinstanz mit bestimmter Art GrösSe durch die beiden Heuristiken sortiert wurde.

6. Laufzeiten der Heuristiken

Es werden je Art und GrösSe der Instanz 3 Messungen durchgeführt, von denen am Ende der Durchschnitt berechnet wird. Die Einheiten der Messungen sind jeweils Sekunden (s). Die Tabellen der Laufzeitenmessungen sind später im **Anhang B** aufgeführt.

A. Ergebnis-Qualität

A.1. Abstandssummen der gemischten Bilder

Messlauf \ Art	#1	#2	#3
g1	834.302888687	763.215338135	915.014689799
g2	523.503612176	523.503612176	573.212169601
p1	715.996547763	715.996547763	713.029264826
p2	428.438632521	388.942148901	406.337541942
t1	248.472480733	263.107387688	250.725344778
t2	974.582228968	885.182192758	950.296186762

Tabelle 1: Abstandssummen der gemischten Instanzen der GrösSe 10

Messlauf \ Art	#1	#2	#3
g1	3386.10358879	2756.28773204	3230.96412561
g2	2168.00156018	2255.88720735	2185.99466697
p1	7753.696801	7839.42728232	7740.21345487
p2	6352.0804647	6476.84613688	6320.38864576
t1	3458.82782244	3465.62528179	3457.03636612
t2	12478.3784309	12235.0266878	12234.7514628

Tabelle 2: Abstandssummen der gemischten Instanzen der GrösSe 50

Messlauf \ Art	#1	#2	#3
g1	6540.14581498	5656.20209467	6638.07923591
g2	3054.30639553	3074.15843039	3139.57575107
p1	20765.541476	20933.9591893	20762.0657001
p2	18215.8892259	17769.2556008	18113.899511
t1	12254.7534816	12363.7166301	12333.0075423
t2	20995.3984479	21076.5563615	20986.0028343

Tabelle 3: Abstandssummen der gemischten Instanzen der GrösSe 100

A.2. Abstandssummen der sortierten Bilder

A.2.1. Heuristik 1

Messlauf \ Art	#1	#2	#3	\emptyset
g1				
g2				
p1				
p2				
t1				
t2				

Tabelle 4: Heuristik 1: Abstandssummen der sortierten Instanzen der GrösSe 10

Art \ Messlauf	#1	#2	#3	\emptyset
g1				
g2				
p1				
p2				
t1				
t2				

Tabelle 5: Heuristik 1: Abstandssummen der sortierten Instanzen der GrösSe 50

Art \ Messlauf	#1	#2	#3	\emptyset
g1				
g2				
p1				
p2				
t1				
t2				

Tabelle 6: Heuristik 1: Abstandssummen der sortierten Instanzen der GrösSe 100

A.2.2. Heuristik 2

Art \ Messlauf	#1	#2	#3	\emptyset
g1				
g2				
p1				
p2				
t1				
t2				

Tabelle 7: Heuristik 2: Abstandssummen der sortierten Instanzen der GrösSe 10

Art \ Messlauf	#1	#2	#3	\emptyset
g1				
g2				
p1				
p2				
t1				
t2				

Tabelle 8: Heuristik 2: Abstandssummen der sortierten Instanzen der GrösSe 50

Art	Messlauf	#1	#2	#3	ϕ
	g1				
	g2				
	p1				
	p2				
	t1				
	t2				

Tabelle 9: Heuristik 2: Abstandssummen der sortierten Instanzen der GrösSe 100

B. Laufzeiten der Heuristiken

B.0.1. Heuristik 1

Art \ Messlauf	#1	#2	#3	ø
g1				
g2				
p1				
p2				
t1				
t2				

Tabelle 10: Heuristik 1: Laufzeiten bei Instanzen der Grösse 10

Art \ Messlauf	#1	#2	#3	ø
g1				
g2				
p1				
p2				
t1				
t2				

Tabelle 11: Heuristik 1: Laufzeiten bei Instanzen der Grösse 50

Art \ Messlauf	#1	#2	#3	ø
g1				
g2				
p1				
p2				
t1				
t2				

Tabelle 12: Heuristik 1: Laufzeiten bei Instanzen der Grösse 100

B.0.2. Heuristik 2

Art \ Messlauf	#1	#2	#3	ø
g1				
g2				
p1				
p2				
t1				
t2				

Tabelle 13: Heuristik 2: Laufzeiten bei Instanzen der Grösse 10

Art \ Messlauf	#1	#2	#3	ø
g1				
g2				
p1				
p2				
t1				
t2				

Tabelle 14: Heuristik 2: Laufzeiten bei Instanzen der Grösse 50

Art	Messlauf	#1	#2	#3	ϕ
	g1				
	g2				
	p1				
	p2				
	t1				
	t2				

Tabelle 15: Heuristik 2: Laufzeiten bei Instanzen der Grösse 100

Tabellenverzeichnis

1.	Abstandssummen der gemischten Instanzen der Größe 10	3
2.	Abstandssummen der gemischten Instanzen der Größe 50	3
3.	Abstandssummen der gemischten Instanzen der Größe 100	3
4.	Heuristik 1: Abstandssummen der sortierten Instanzen der Größe 10	3
5.	Heuristik 1: Abstandssummen der sortierten Instanzen der Größe 50	4
6.	Heuristik 1: Abstandssummen der sortierten Instanzen der Größe 100	4
7.	Heuristik 2: Abstandssummen der sortierten Instanzen der Größe 10	4
8.	Heuristik 2: Abstandssummen der sortierten Instanzen der Größe 50	4
9.	Heuristik 2: Abstandssummen der sortierten Instanzen der Größe 100	5
10.	Heuristik 1: Laufzeiten bei Instanzen der Größe 10	6
11.	Heuristik 1: Laufzeiten bei Instanzen der Größe 50	6
12.	Heuristik 1: Laufzeiten bei Instanzen der Größe 100	6
13.	Heuristik 2: Laufzeiten bei Instanzen der Größe 10	6
14.	Heuristik 2: Laufzeiten bei Instanzen der Größe 50	6
15.	Heuristik 2: Laufzeiten bei Instanzen der Größe 100	7

Literatur

- [BGJ01] Ziv Bar-Joseph, David K Gifford und Tommi S Jaakkola.
„Fast optimal leaf ordering for hierarchical clustering“.
In: *Bioinformatics* 17.suppl_1 (2001), S22–S29.
URL: <https://watermark.silverchair.com/17S022.pdf> (siehe S. 1).