

Übung Algorithm Engineering

Toni Barth und Max Haarbach

24. September 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	511,890729981	593,989925231	593,989925231	566,6235268143
g2	378,273825497	249,873800568	297,765601491	308,6377425187
p1	579,379823257	522,444733052	536,133377774	545,9859780277
p2	391,556113931	401,822571285	396,699104033	396,6925964163
t1	207,463887593	192,909859229	194,893376625	198,4223744823
t2	838,263047345	869,673011642	838,80454995	848,9135363123

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

Messlauf \ Art	#1	#2	#3	ϕ
g1	2736,97257701	2551,29814308	2888,04580254	2725,438840877
g2	1596,42449569	1486,68581038	1344,80513322	1475,971813097
p1	6963,39812615	6820,63085512	6940,72230471	6908,25042866
p2	6060,36325688	5988,33180827	6031,321592	6026,67221905
t1	2834,37185044	3058,84209343	2912,4480459	2935,220663257
t2	10838,6803098	10993,4434053	11082,3927782	10971,50549777

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

Messlauf \ Art	#1	#2	#3	ϕ
g1	4835,82352844	4756,40532119	4441,21981118	4677,81622027
g2	2214,59876015	2293,71637533	2352,17062442	2286,82858663
p1	20010,8195814	19889,8369096	20041,5197608	19980,72541727
p2	17071,5189857	16979,5694001	16875,7776209	16975,62200223
t1	10585,1207027	10728,5233067	10885,4157573	10733,01992223
t2	20173,4030102	20210,8315004	19582,1992343	19988,8112483

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

A.2.2. Heuristik 2

Messlauf \ Art	#1	#2	#3	ϕ
g1	426,752089999	426,752089999	426,752089999	426,752089999
g2	249,873800568	249,873800568	249,873800568	249,873800568
p1	522,444733052	522,444733052	522,444733052	522,444733052
p2	373,400786545	373,400786545	373,400786545	373,400786545
t1	192,909859229	192,909859229	192,909859229	192,909859229
t2	838,263047345	838,263047345	838,263047345	838,263047345

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

Messlauf \ Art	#1	#2	#3	ϕ
g1	1715,86645274	1085,82959774	1085,82959774	1295,84188274
g2	604,636533543	604,636533543	604,636533543	604,636533543
p1	6587,51285829	6587,51285829	6587,51285829	6587,51285829
p2	5492,5290117	5492,5290117	5492,5290117	5492,5290117
t1	2798,7771451	2799,35407759	2798,7771451	2798,96945593
t2	10183,381895	10183,381895	10183,381895	10183,381895

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

Messlauf Art	#1	#2	#3	\varnothing
g1	1892,391478	1923,40968002	1892,391478	1902,730878673
g2	990,337428662	1089,07985883	1137,06979326	1072,1623602507
p1	18371,6242832	18371,6242832	18371,6242832	18371,6242832
p2	14399,1155256	14399,1155256	14399,1155256	14399,1155256
t1	9955,00625808	9955,00625808	9955,00625808	9955,00625808
t2	18823,7799399	18823,7799399	18823,7799399	18823,7799399

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	ϕ
g1	0,002	0,002	0,002	0,002
g2	0,002	0,002	0,002	0,002
p1	0,002	0,002	0,002	0,002
p2	0,002	0,002	0,002	0,002
t1	0,002	0,002	0,002	0,002
t2	0,002	0,002	0,002	0,002

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

Art \ Messlauf	#1	#2	#3	ϕ
g1	0,044	0,042	0,042	0,0427
g2	0,042	0,042	0,042	0,042
p1	0,043	0,042	0,044	0,043
p2	0,043	0,044	0,042	0,043
t1	0,043	0,043	0,042	0,0423
t2	0,043	0,043	0,043	0,043

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

Art \ Messlauf	#1	#2	#3	ϕ
g1	0,176	0,175	0,176	0,1757
g2	0,175	0,174	0,174	0,1747
p1	0,184	0,177	0,177	0,1793
p2	0,175	0,180	0,173	0,176
t1	0,175	0,176	0,175	0,1753
t2	0,176	0,180	0,174	0,1767

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

B.0.2. Heuristik 2

Art \ Messlauf	#1	#2	#3	ϕ
g1	0,001	0,001	0,001	0,001
g2	0,001	0,001	0,001	0,001
p1	0,001	0,001	0,001	0,001
p2	0,001	0,001	0,001	0,001
t1	0,001	0,001	0,001	0,001
t2	0,001	0,001	0,001	0,001

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

Messlauf Art	#1	#2	#3	ϕ
g1	0,327	0,336	0,322	0,3283
g2	0,638	0,668	0,661	0,6557
p1	0,136	0,135	0,140	0,137
p2	0,125	0,127	0,126	0,126
t1	0,381	0,373	0,389	0,381
t2	0,166	0,166	0,167	0,1667

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

Messlauf Art	#1	#2	#3	ϕ
g1	10,436	13,532	10,244	11,404
g2	14,771	14,677	14,596	14,6813
p1	5,441	5,334	5,342	5,3723
p2	22,941	22,983	22,961	22,9617
t1	8,215	7,448	7,480	7,7143
t2	12,111	12,275	12,146	12,1773

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).