



# On the influence of task size and template provision on solution similarity

Tobias Haan, Michael Striewe

# Motivation

- Zwischen Lösungen von Programmieraufgaben kann es viele **Unterschiede** geben:
  - In der Strategie (z. B. rekursiv vs. iterativ)
  - In der Konstruktion (z. B. for-Schleife vs. while-Schleife)
  - In der Syntax (z. B. `i++` vs. `i=i+1`)
  - In der Namensgebung für Identifizierer
- Gleichwohl gibt es viele Werkzeuge und Einsatzbereiche, die auf **Ähnlichkeiten** basieren
  - Automatische Bewertung [Na07, TRB04]
  - Semi-Automatische Unterstützung bei der Bewertung [He17]
  - Bestimmung der (relativen) Aufgabenschwierigkeit [SG13]
  - Suche nach Plagiaten [PPM00]
- **Aber was ist eigentlich unser Bezugspunkt? Wann ist eine Lösung „ähnlich“ oder „unterschiedlich“?**

# Forschungsfrage und Vorgehen

- Es lassen sich schnell **grobe Hypothesen oder Fragen** finden:
  - Wie hängt die (durchschnittliche) Ähnlichkeit der Lösungen mit der (durchschnittlichen) Größe der Lösungen zusammen?
  - Welchen Einfluss haben vorgegebene Lösungstemplates?
  - Sind korrekte Lösungen ähnlicher zueinander?
  - Stimmen verschiedene Ähnlichkeitsmaße überein?
- Vorgehen zur **explorativen Näherung an präzisere Hypothesen**:
  - Sechs Aufgaben aus einer Einführungsvorlesung zur Programmierung in Java; jeweils mit mehreren Hundert Lösungen
  - Drei Verfahren/Werkzeuge zur Berechnung von Ähnlichkeiten
    - Clone Doctor
    - Deckard
    - LAV
  - Ergebnisse normalisieren
  - Mittelwert und Median vergleichen

# Datenbasis

	Durchschnittliche Länge der Lösung (LOC)	Länge des Templates (LOC)	Anteil des Templates an der durchschnittlichen Lösung	Gesamtzahl an Lösungen	Anzahl korrekter Lösungen
Aufgabe 1 Variablen und if	45	15	0,33	1352	452
Aufgabe 2 Konstruktoren, Arrays, Schleifen	228	145	0,64	4058	1182
Aufgabe 3 Datenstruktur Liste	158	46	0,29	3078	795
Aufgabe 4 Datenstruktur Binärer Suchbaum	529	365	0,69	4790	652
Aufgabe 5 Vererbung	69	34	0,49	5422	1897
Aufgabe 6 Enumerations, Interfaces, Maps	211	116	0,55	3432	1005

# Ähnlichkeitswerte

	Clone Doctor		Deckard		LAV		Mittelwert	Median
	Alle	Nur korrekte	Alle	Nur korrekte	Alle	Nur korrekte		
Aufgabe 1	0,480	0,448	0,224 0,196	0,221 0,195	0,238 0,209	0,235 0,207		
Aufgabe 2	0,744	0,743	0,447 0,447	0,467 0,468	0,442 0,442	0,462 0,463		
Aufgabe 3	0,713	0,695	0,196 0,156	0,195 0,174	0,207 0,166	0,207 0,186		
Aufgabe 4	0,830	0,817	0,386 0,370	0,366 0,360	0,408 0,394	0,389 0,383		
Aufgabe 5	0,759	0,842	0,398 0,394	0,454 0,454	0,422 0,421	0,481 0,486		
Aufgabe 6	0,799	0,779	0,367 0,337	0,422 0,420	0,388 0,359	0,448 0,449		

# Ähnlichkeit

Werte von Clone Doctor höher als von Deckard und LAV

	Clone Doctor		Deckard		LAV	
	Alle	Nur korrekte	Alle	Nur korrekte	Alle	Nur korrekte
Aufgabe 1	0,480	0,448	0,224 0,196	0,221 0,195	0,238 0,209	0,235 0,207
Aufgabe 2	0,744	0,743	0,447 0,447	0,467 0,468	0,442 0,442	0,462 0,463
Aufgabe 3	0,713	0,695	0,196 0,156	0,195 0,174	0,207 0,166	0,207 0,186
Aufgabe 4	0,830	0,817	0,386 0,370	0,366 0,360	0,408 0,394	0,389 0,383
Aufgabe 5	0,759	0,842	0,398 0,394	0,454 0,454	0,422 0,421	0,481 0,486
Aufgabe 6	0,799	0,779	0,367 0,337	0,422 0,420	0,388 0,359	0,448 0,449

Kein klarer Trend, ob korrekte Lösungen höhere Ähnlichkeit haben

# Ähnlichkeitswerte

	Clone Doctor		Deckard		LAV	
	Alle	Nur korrekte	Alle	Nur korrekte	Alle	Nur korrekte
Aufgabe 1	0,480	0,448	0,224	0,221	0,238	0,235
			0,196	0,195	0,209	0,207
Aufgabe 2	0,744	0,743	0,447	0,467	0,442	0,462
			0,447	0,468	0,442	0,463
Aufgabe 3	0,713	0,695	0,196	0,195	0,207	0,207
			0,156	0,174	0,166	0,186
Aufgabe 4	0,830	0,817	0,386	0,366	0,408	0,389
			0,370	0,360	0,394	0,383
		0,842	0,398	0,454	0,422	0,481
			0,394	0,454	0,421	0,486
		0,779	0,367	0,422	0,388	0,448
			0,337	0,420	0,359	0,449

Unterschied zwischen allen Lösungen und nur korrekten >0.05; sonst fast immer <0.02

# Ähnlichkeitswerte

	Clone Doctor		Deckard		LAV	
	Alle	Nur korrekte	Alle	Nur korrekte	Alle	Nur korrekte
Aufgabe 1	0,480	0,448	0,224 0,196	0,221 0,195	0,238 0,209	0,235 0,207
Aufgabe 2	0,744	0,743	0,447 0,447	0,467 0,468	0,442 0,442	0,462 0,463
Aufgabe 3			0,196 0,156	0,195 0,174	0,207 0,166	0,207 0,186
Aufgabe 4			0,386 0,370	0,366 0,360	0,408 0,394	0,389 0,383
Aufgabe 5			0,398 0,394	0,454 0,454	0,422 0,421	0,481 0,486
Aufgabe 6			0,367 0,337	0,422 0,420	0,388 0,359	0,448 0,449

Kleiner Code-Anteil des Templates (0,29) und geringste Ähnlichkeitswerte

(Zweit-)Höchste Ähnlichkeitswerte, aber nur mittlerer Code-Anteil des Templates (0,49)

# Ähnlichkeitswerte

	Clone Doctor		Deckard		LAV	
	Alle	Nur korrekte	Alle	Nur korrekte	Alle	Nur korrekte
Aufgabe 1	0,480	0,448	0,224 0,196	0,221 0,195	0,238 0,209	0,235 0,207
Aufgabe 2	0,744	0,743	0,447 0,447	0,467 0,468	0,442 0,442	0,462 0,463
Aufgabe 3	0,713	0,695				
Aufgabe 4	0,830	0,817	0,386 0,370	0,366 0,360	0,408 0,394	0,389 0,383
Aufgabe 5	0,759	0,842	0,398 0,394	0,454 0,454	0,422 0,421	0,481 0,486
Aufgabe 6	0,799	0,779	0,367 0,337	0,422 0,420	0,388 0,359	0,448 0,449

Fast gleicher Anteil des Code-Templates (0,64 vs. 0,69), aber abweichende Ähnlichkeiten

# Ähnlichkeitswerte

	Clone Doctor		Deckard		LAV	
	Alle	Nur korrekte	Alle	Nur korrekte	Alle	Nur korrekte
Aufgabe 1	0,480	0,448	0,224 0,196	0,221 0,195	0,238 0,209	0,235 0,207
Aufgabe 2	0,744	0,743	0,447 0,447	0,467	0,442	0,462 0,463
Aufgabe 3	0,713	0,695	0,196 0,156			0,207 0,186
Aufgabe 4	0,830	0,817	0,386 0,370	0,360	0,400 0,394	0,389 0,383
Aufgabe 5	0,759	0,842	0,398 0,394	0,454 0,454	0,422 0,421	0,481 0,486
Aufgabe 6	0,799	0,779	0,367 0,337	0,422 0,420	0,388 0,359	0,448 0,449

30 bzw. 35 LOC von Studierenden, aber völlig unterschiedliche Ähnlichkeitswerte

# Ähnlichkeitswerte

	Clone Doctor		Deckard		LAV	
	Alle	Nur korrekte	Alle	Nur korrekte	Alle	Nur korrekte
Aufgabe 1	0,480	0,448	0,224 0,196	0,221 0,195	0,238 0,209	0,235 0,207
Aufgabe 2	0,744	0,743	0,447 0,447	0,467 0,468	0,442 0,442	0,462 0,463
Aufgabe 3	0,713	0,695	0,196 0,156	0,195 0,174	0,207 0,166	0,207 0,186
Aufgabe 4	0,777	0,777	0,386 0,370	0,366 0,360	0,408 0,394	0,389 0,383
Aufgabe 5	0,774	0,774	0,398 0,394	0,454 0,454	0,422 0,421	0,481 0,486
Aufgabe 6	0,799	0,779	0,367 0,337	0,422 0,420	0,368 0,359	0,448 0,449

Ähnliche Werte für alle, aber große Unterschiede bei den korrekten Lösungen. Liegt es an 164 vs. 35 LOC Beitrag von Studierenden?

7

4

# Diskussion

- Fehlende offensichtliche Korrelationen nicht gänzlich überraschend
  - Es gibt sicher noch **andere Faktoren**
  - Aber: Aufgabe 1 mit sehr geringer Ähnlichkeit, obwohl klein und mit sehr engem Scope
  - Und Aufgabe 2 gleich mit viel höhere Ähnlichkeit
- Zwei mögliche Faktoren für die **Unterschiede zwischen allen und korrekten Lösungen**:
  - Korrekte Lösungen folgen wenigen Mustern; Abweichungen führen zu Fehlern -> Ähnlichkeit für alle Lösungen geringer als für korrekte
  - Fehlerhafte Lösungen könnten unvollständig sein; korrekte Lösungen haben höheren Eigenanteil -> Ähnlichkeit für falsche Lösungen höher als für korrekte

# Fazit und nächste Schritte

- Einfache Prediktoren für die Ähnlichkeit gibt es nicht
  - **Wir haben noch immer keinen Bezugspunkt**
  - Zumindest gibt es Ideen für Hypothesen
- Naheliegende Maßnahmen:
  - Mehr **Kategorien bilden**
    - Geht im Kontext automatische Bewertung natürlich besonders gut auf Basis von Ergebnissen
    - Unvollständige Lösungen als eigene Kategorie
  - Andere Faktoren mit berücksichtigen
    - Aber wie quantifiziert man die Vorgaben aus der Aufgabenstellung?
  - Cluster suchen und interpretieren
- Aber auch: Entspricht die **automatische Messung der Ähnlichkeit** überhaupt dem Verständnis aus der Lehre?