

Genomics

Freie Universität Berlin, Institut für Informatik

Peter Robinson

Wintersemester 2015/2016

1. Übungsblatt vom 15. Oktober 2015

Diskussion am 22.10.2015

Die folgenden Aufgaben basieren locker auf dem informatischen Evolutionsmodell von Richard Dawkins (1986). Die praktische Aufgabe stellt eine Art einfachen genetischen Algorithmus dar. Zwei Tipps:

1. Wenn es schwierig ist, die Wahrscheinlichkeit eines Ereignisses zu berechnen, probieren Sie, die Wahrscheinlichkeit zu berechnen, dass das Ereignis nicht eintritt
2. Hier ist eine nützliche Approximation: $(1 - u)^n \approx e^{-un}$ für u klein und n groß

Wir wollen die "Evolution" eines Spruches simulieren

"Arbeite klug, nicht hart."

Der Einfachheit halber werden wir den Spruch so darstellen

ARBEITEKLUGNICHTHART

Unser Spruch enthält also 20 Buchstaben. Wir lassen diesen Spruch als evolutionäre "optimal" gelten und untersuchen die Evolution des Spruchs aus einem Alphabet mit 26 Buchstaben zufälligen gewählten Wort mit 20 Buchstaben (der Einfachheit halber wollen wir nur ASCII-Zeichen verwenden, also kein ä,ö,ü oder ß).

Aufgabe 1.

Wieviele Wörter mit 20 Buchstaben gibt es bei einem Alphabet mit 26 Buchstaben?

Aufgabe 2.

a) Was ist die Wahrscheinlichkeit, dass beim zufällig gewählten Wort mindestens ein Buchstabe korrekt ist? Nehmen Sie an, dass die Wahrscheinlichkeit für jeden einzelnen Buchstaben, korrekt zu sein, $\frac{1}{26}$ beträgt.

b) Was ist die Wahrscheinlichkeit, dass beim zufällig gewählten Wort genau ein Buchstabe korrekt ist?

Aufgabe 3.

Zehn Kopien des ursprünglichen zufällig erzeugten Wortes werden gemacht. Bei jedem Kopiervorgang gibt es eine Wahrscheinlichkeit von 99/100, dass ein Buchstabe korrekt kopiert wird und

1/100, dass er falsch kopiert wird (eine "Mutation" – mit einer gleichen Wahrscheinlichkeit zu einem beliebigen Buchstaben unter den 25 anderen Buchstaben).

Nehmen Sie an, dass beim ursprünglichen Wort kein Buchstabe korrekt war. Wie hoch ist die Wahrscheinlichkeit, dass bei mindestens einer der zehn Kopien mindestens ein Buchstabe korrekt ist?

Aufgabe 4.

Nun simulieren wir die Evolution wie folgt. Die beste unter den 10 Kopien (die Kopien mit der höchsten Anzahl von passenden Buchstaben) wird als "Vorfahr" der nächsten Generation gewählt. Falls mehrere Kopien gleich gut sind, wird eine zufällig ausgesucht. Die beste Kopien wird dann verwendet, um zehn neue Kopien wie oben zu erzeugen.

Nehmen Sie nochmals an, dass beim ursprünglichen Wort kein Buchstabe korrekt war. Wie viele Generationen werden ungefähr (in Erwartung) notwendig sein, bis ein Wort mit mindestens einem korrektne Buchstaben vorhanden ist?

Aufgabe 5. (Optional)

Nun schreiben Sie ein Computerprogramm, um die Evolution des Wortes ARBEITEKLUGNICHTHART zu simulieren. Berichten Sie über das Ergebnis. Sie dürfen die Skriptsprache oder kompilierte Sprache Ihrer Wahl verwenden. Ich habe ein Programm geschrieben, dessen Ausgabe wie folgt aussieht (Format lautet generation:word [score], eine Ausgabe erfolgt nur bei einer Verbesserung des scores):

```
1:HDWEWUTEKRCNEZYTDOMR [3]
7:HDBEWUTEJRCNEZYTDOMR [4]
34:HQB EWUTEJGJNEZYTHOWW [5]
75:HLBEKUTEIGJNEZYTHCWT [6]
148:RXBEKULEVQANE EYTHAMT [7]
162:RXBEKULEVQANE EHTAMT [8]
163:RRBEKULEVQANE EHTAMT [9]
215:VRBEARLQVQANE HHTHART [10]
233:VRBEARLWVQANE CHTHART [11]
247:VRBEARLKVQANE CHTHART [12]
261:WRBEIRLKPQANE CHTHART [13]
301:WRBEIRLKVQGNE CHTHART [14]
375:IRBEIDMKBQGNICHTHART [15]
377:IRBEIDMKLQGNICHTHART [16]
439:ARBEIEMKLIGNICHTHART [17]
475:ARBEIEEKLIGNICHTHART [18]
791:ARBEIOEKLUGNICHTHART [19]
826:ARBEITEKLUGNICHTHART [20]
```