Universität des Saarlandes
FR Informatik

Kurt Mehlhorn und Antonios Antoniadis **WiSe 2019/20**

Übungen zu Ideen der Informatik

https://www.mpi-inf.mpg.de/departments/algorithms-complexity/teaching/winter19/ideen/

Blatt 8 Abgabeschluss: 9.12.2019

Aufgabe 1 (*15 Punkte*) Editierdistanz: Wie schwer ist es ein Wort in ein anderes zu überführen. Wir dürfen dabei Buchstaben streichen, Buchstaben einfügen und einen Buchstaben durch einen anderen Buchstaben ersetzen. Jede Operation hat dabei kosten 1.
Man kann etwa MOTTE in LOTTE überführen, in dem man das M durch ein L ersetzt. Die Kosten sind 1.
Man kann OTTO in TOTO mit Kosten zwei überführen: durch Ersetzen der ersten beiden Buchstaben oder durch Einfügen eines Ts am Anfang und durch Streichen eines der beiden Ts.

O	T	T	O	oder		O	T	T	O
T	O	T	O		T	O		T	O

a) Sie haben sicher schon gelesen, dass die Genome von Menschen und Menschenaffen zu 99,4% übereinstimmen. Was könnte der Zusammenhang zwischen dieser Aussage und dieser Übung sein?

Seien unsere beiden Worte nun $x = x_1x_2 \dots x_n$ und $y = y_1y_2 \dots y_m$. Dabei steht jedes x_i und y_j für einen Buchstaben. Für das Wort MOTTE ist $n = 5$ und $x_1 = M$, $x_2 = O$, $x_3 = T$, $x_4 = T$ und $x_5 = E$. Mit $x_1 \dots x_i$ bezeichnen wir den Präfix (das Anfangswort) von x der Länge i. In unserem Beispiel ist $x_1x_2x_3 = MOT$. Etwas ungewohnt sprechen wir auch vom leeren Präfix. Das ist der Präfix aus null Buchstaben.
Wir füllen nun eine rechteckige Tabelle T mit $n+1$ Zeilen und $m+1$ Spalten aus. An der Stelle $T[i, j]$ (Eintrag in der i-ten Zeile und j-ten Spalte) wollen wir niedrigsten Kosten für die Überführung von $x_1 \dots x_i$ in $y_1 \dots y_j$ eintragen. Dabei läuft i von 0 bis n und j von 0 bis m. In $T[n, m]$ stehen dann die Kosten für die Überführung von x in y.
Ich erkläre jetzt, wie wir die nullte Zeile und Spalte einfüllen. Wir initialisieren $T[0, 0] = 0$, $T[0, j] = j$ für $1 \le j \le m$ und $T[i, 0] = i$ für $1 \le i \le n$. In $T[0, 0]$ steht eine Null, da das gerade die Kosten der Überführung des leeren Präfixes von x (bestehend aus null Buchstaben) in den leeren Präfix von y sind. In $T[0, j]$ steht j weil das die Kosten der Überführung von $y_1 \dots y_j$ in den leeren Präfix von x (der Präfix aus null Buchstaben) sind. Man muss alle j Buchstaben streichen um das leere Worte zu bekommen.
Für $i \ge 1$ und $j \ge 1$ berechnen wir

$$T[i, j] = \begin{cases} \min(1 + T[i-1, j], 1 + T[i, j-1], T[i-1, j-1]) & \text{if } x_i = y_j \\ \min(1 + T[i-1, j], 1 + T[i, j-1], 1 + T[i-1, j-1]) & \text{if } x_i \neq y_j \end{cases}$$

Für OTTO und TOTO erhalten wir folgende Tabelle mit 5 Spalten und Zeilen.

		O	T	T	O
	0	1	2	3	4
T	1	1	1	2	3
O	2	1	2	2	2
T	3	2	2	2	3
O	4	3	3	3	2

a) Verifizieren und gegebenfalls korrigieren sie die obige Tabelle und geben sie für jedes Kästchen $T[i, j]$ mit $i \ge 1$ und $j \ge 1$ durch einen Pfeil an, durch welches andere Kästchen sein Wert bestimmt wurde. Wenn es mehrere Möglichkeiten gibt, zeichnen sie alle Möglichkeiten ein.

b) Stellen sie die Tabelle für LOTTA und MOTTE auf.

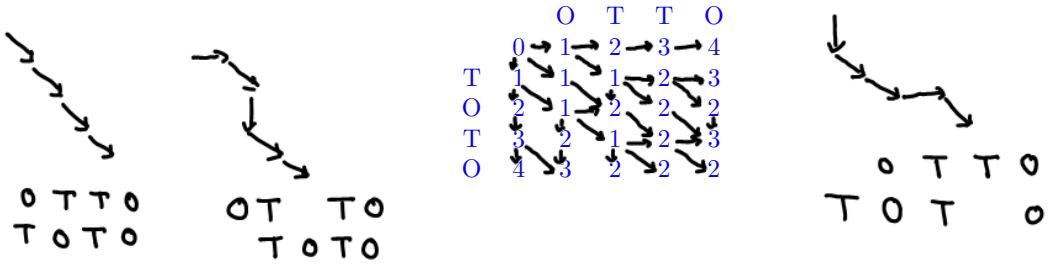
c) Wir kann man aus den Tabellen ablesen, wie man das eine Wort möglichst günstig in das andere Wort überführen kann.

d) Geben sie eine Begründung für die Definition von $T[i,j]$ an. Etwa so. Ich kann $x_1 \dots x_i$ in $y_1 \dots y_j$ überführen, indem ich $x_1 \dots x_{i-1}$ in $y_1 \dots y_j$ überführe und x_i streiche oder UND NUN FAHREN SIE FORT.

Lösung:

a) Das menschliche Genom ist ein Wort über den 4 Buchstaben des genetischen Codes. 99,4% Übereinstimmung bedeutet, dass sie nur den Bruchteil 6/1000 der Buchstaben ändern müssen, um aus der menschlichen DNA eine Menschenaffen DNA zu machen. Die Editierdistanz ist also nur 6/1000 der Länge des Genoms.

b) Hier ist die Tabelle mit den Pfeilen.



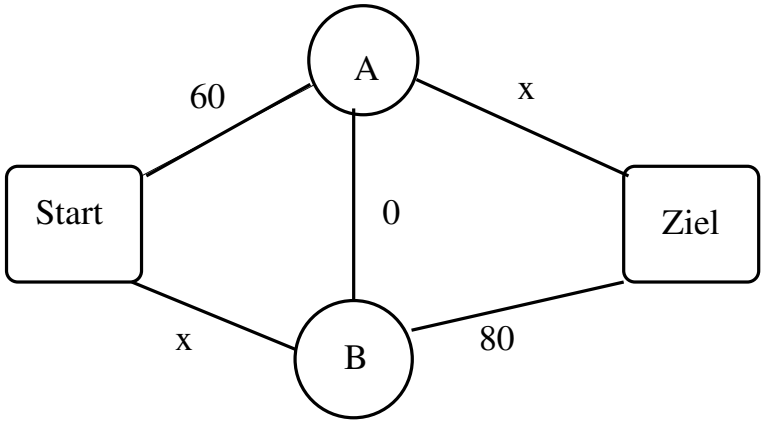
c)

		L	O	T	T	A
	0	1	2	3	4	5
M	1	1	2	3	4	5
O	2	2	1	2	3	4
T	3	3	2	1	2	3
T	4	4	3	2	1	2
E	5	5	4	3	2	2

d) Wir beginnen im Kästchen $T[n,m]$ und verfolgen die Pfeile zurück bis wir in $[0,0]$ ankommen. Im Beispiel gibt es mehrere Wege zurück.

e) Ich kann $x_1 \dots x_i$ in $y_1 \dots y_j$ überführen, indem ich $x_1 \dots x_{i-1}$ in $y_1 \dots y_j$ überführe und x_i streiche oder indem ich $x_1 \dots x_i$ in $y_1 \dots y_{j-1}$ überführe und y_j streiche oder indem ich $x_1 \dots x_{i-1}$ in $y_1 \dots y_{j-1}$ überführe und x_i in y_j überführe. Falls $x_i = y_j$ entstehen dafür keine Kosten.

Aufgabe 2 (*15 Punkte*)



100 Autos wollen von Start nach Ziel fahren. Die Fahrzeiten sind wie angegeben. Auf der Straße von Start nach B und von A nach Ziel ist die Fahrzeit x Minuten, wenn sie von x Autos befahren wird. Nehmen Sie zuächst an, dass die Straße zwischen A und B NICHT existiert.

a) Was ist das globale Optimum (Globales Optimum = minimale Gesamtfahrzeit)? Wie viele Autos fahren oben rum und wieviele Fahren unten rum? Was sind die Fahrzeiten für die einzelnen Fahrer. Stellt sich dieses Optimum auch ein, wenn jeder einzelne Fahrer seine Fahrzeit optimiert?

b) Wir nehmen nun die Straße zwischen A und B hinzu. Was ist nun das globale Optimum? Welches Gleichgewicht stellt sich ein, wenn jeder Fahrer seine Fahrzeit optimiert? Wie sind die Fahrzeiten für die einzelnen Fahrer im sozialen Optimum.

Lösung:

a) Wir nehmen zunächst an, dass es keine Straße zwischen A und B gibt.

Wenn x Fahrer oben rum fahren und $100-$ unten rum, dann ist die Gesamtfahrzeit

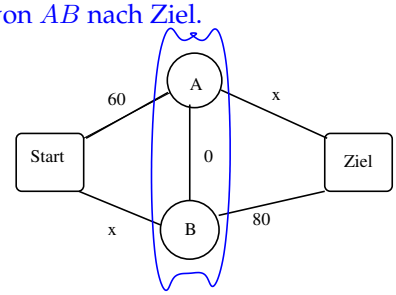
$$x \cdot (60 + x) + (100 - x) \cdot (80 + (100 - x)) = 60x + x^2 + 18000 - 80x + x^2 = 18000 - 220x + 2x^2.$$

Ableiten zeigt, dass das Minimum für $n = 55$ angenommen wird. Die oben-Fahrer brauchen 115 Minuten und die unten-Fahrer brauchen 125 Minuten. Die Gesamtfahrzeit ist 11950 Minuten.

Wenn jeder Fahrer selbst entscheidet, stellt sich ein Gleichgewicht stellt ein, in dem die unten-Fahrer genauso lang brauchen wie die oben-Fahrer.

Wenn x Personen oben fahren, ist ihre Fahrzeit $60 + x$. Die Fahrzeit der unten-Fahrer ist $180 - x$. Gleichlange Fahrzeit haben wir für $60+x = 180-x$, also $x = 60$. Dann ist die Gesamtfahrzeit $100 \cdot 120 = 12000$.

b) Jetzt kommt die kostenlose Verbindung zwischen A und B dazu. Das entspricht einem Verschmelzen der Knoten A und B wie in der Abbildung gezeigt. Wir haben damit zwei Straßen von Start zu dem Knoten AB und zwei Straßen von AB nach Ziel.



Nash-Gleichgewicht: In der linken Hälfte fahren 60 unten und 40 oben. Dann ist die Fahrzeit für jeden 60 Minuten. In der rechten Hälfte fahren 80 oben und 20 unten. Dann ist die Fahrzeit für jeden 140 Minuten. Die Gesamtfahrzeit ist 14000 Minuten. Das ist schlechter als vor dem Bau der Straße. 40 Autos wechseln von B nach A.

Soziales Optimum: Wenn x Autos unten herum von Start nach AB fahren dann ist die Gesamtfahrzeit $x \cdot x + 60(100 - x) = 6000 - 60x + x^2$. Die Ableitung ist Null für $x = 30$. Also fahren 30 Autos über die Straße Start – B und 70 über die Straße Start – A. Die Gesamtfahrzeit für das linke Teilnetzwerk ist $30 \cdot 30 + 60 \cdot 70 = 5100$.

Nun zum rechten Teilnetzwerk. Wenn x Autos die Straße A – Ziel benutzen, dann ist die Gesamtfahrzeit $x^2 + 80(100 - x) = 8000 - 80x + x^2$. Die Ableitung ist Null für $x = 40$. Also fahren 40 Autos auf der Straße A – Ziel und 60 Autos auf der Straße B – Ziel. Die Gesamtfahrzeit für das rechte Teilnetzwerk ist $40 \cdot 40 + 60 \cdot 80 = 6400$.

Die Gesamtfahrzeit ist $5100 + 6400 = 11500$. Das ist besser als vor dem Bau der Straße. 30 Autos wechseln von A nach B.

Die Fahrzeiten sind stark unterschiedlich: 30 fahren Start – B – Ziel mit einer Gesamtfahrzeit von 110 Minuten, 30 fahren Start – A – B – Ziel mit einer Fahrzeit von 140 Minuten, und 40 fahren Start – A – Ziel mit einer Fahrzeit von 100 Minuten.

Auktionen, Gleichgewichte, Nutzenmaximierende Agenten war spannend ☐ okay ☐ langweilig ☐
schwierig ☐ okay ☐ einfach ☐