Aufgaben zum Tut am 30.04.2007

Thomas Pajor

30. April 2007

Aufgabe 1.

Gegeben seien die Zeichenketten $\alpha = \texttt{insel}$ und $\beta = \texttt{senil}$.

(a) Berechnen Sie den Hamming-Abstand zwischen α und β.

(b) Berechnen Sie den Editierabstand mittels dynamischer Programmierung, und geben Sie die Folge der Operationen zu Ihrem Ergebnis an.

Die Kosten für das Löschen seien nun 2.

(b) Beweisen oder widerlegen Sie: δ ist weiterhin eine Metrik.

Lösung.

(a) Der Hamming-Abstand ist definiert als die Anzahl verschiedener Zeichen, wenn man die Zeichenketten übereinander legt. In unserem Fall ergibt sich

$$\begin{array}{ccccc} \text{I} & \text{N} & \text{S} & \text{E} & \text{L} \\ \neq & \neq & \neq & \neq & = \\ \text{S} & \text{E} & \text{N} & \text{I} & \text{L} \end{array}$$

ein Hamming-Abstand von 4, da vier Zeichen nicht übereinstimmen.

(b) Den Editierabstand[1] berechnen wir mittels dynamischer Programmierung. Dies führt zu folgender Tabelle.

Ein Schritt nach unten ($\downarrow$) entspricht dem Löschen, ein Schritt nach rechts ($\rightarrow$) dem Einfügen, und ein diagonaler Schritt ($\searrow$) dem Ersetzen eines Zeichens. Speichert man beim Berechnen jedes Eintrags die durchgeführte Operation, so lässt sich der Pfad vom Ziel zum Startpunkt

[1] In der Literatur oft als Levenshtein-Distanz bezeichnet

		I	N	S	E	L
	0	1	2	3	4	5
S	1	1	2	2	3	4
E	2	2	2	3	2	3
N	3	3	2	3	3	4
I	4	3	3	3	4	4
L	5	4	4	4	4	**4**

zurückverfolgen. In der Tabelle ist dieser Pfad durch die eingerahmten Felder gekennzeichnet. Es ergibt sich also folgende Folge von Operationen:

SENIL → ISENIL → INSENIL ↘ INSENIL ↓ INSEIL ↓ INSEL ↘ INSEL

Achtung! In diesem Falle stimmen der Hamming-Abstand und der Editierabstand überein. Im Allgemeinen ist das nicht so, wenn man beispielsweise die Wörter $\alpha = \texttt{insel}$ und $\beta = \texttt{pinsel}$ vergleicht. Es ist offensichtlich dass der Editierabstand durch Einfügen eines `p` zwischen α und β gerade 1 beträgt. Der Hamming-Abstand ist jedoch 5, wie man an folgender Tabelle sieht:

I	N	S	E	L	
$\neq$	$\neq$	$\neq$	$\neq$	$\neq$	$\neq$
P	I	N	S	E	L

(c) Sind die Kosten für das Löschen eines Zeichens teurer, so verändert sich die Definition der Abstandsfunktion δ wie folgt:

$$\delta(\alpha a, \beta b) := \min \begin{cases} \delta(\alpha, \beta) & \text{falls } a = b \\ \delta(\alpha, \beta) + 1 & \text{falls } a \neq b \\ \delta(\alpha a, \beta) + 2 & \text{Löschen} \\ \delta(\alpha, \beta b) & \text{Einfügen} \end{cases}$$

Diese Funktion ist keine Metrik mehr, da die Symmetrieeigenschaft verletzt ist, wie man sich an einem Gegenbeispiel leicht klarmachen kann. Wählen wir $\alpha = a$ und $\beta = \varepsilon$ (also die Leere Zeichenkette, so ist

$$\delta(a, \epsilon) = 2$$

allerdings ist

$$\delta(\epsilon, a) = 1$$

Es gilt also nicht $\delta(\alpha, \beta) = \delta(\beta, \alpha)$ für alle α, β und somit ist δ keine Metrik.

Aufgabe 2.

Rohöl soll durch ein chemisches Verfahren in Komponenten zerlegt werden:

- schweres Öl S
- mittelschweres Öl M
- leichtes Öl L

Folgende Verfahren stehen zur Verfügung:

10 Einheiten Rohöl ergeben:

- 2 Einheiten S
- 2 Einheiten M
- 1 Einheit L

Kosten: 3€

10 Einheiten Rohöl ergeben:

- 1 Einheit S
- 2 Einheiten M
- 4 Einheiten L

Kosten: 5€

Ein Kunde möchte nun folgende Lieferverpflichtung erfüllt haben:

- 3 Einheiten S
- 5 Einheiten M
- 4 Einheiten L

Sie sollen diesen Auftrag unter Anwendung der beiden Verfahren so kostengünstig wie möglich erfüllen.

(a) Formulieren Sie das Problem als lineares Programm

(b) Bringen Sie das lineare Programm in Standardform

(c) Lösen Sie das LP mit Hilfe des geometrischen Simplexvefahrens

Lösung.

(a) Wir definieren uns für jedes Verfahren eine Variable x_i. x_1 stehe für die Anwendung des ersten Verfahrens und x_2 für die Anwendung des zweiten Verfahrens. Somit lautet die zu minimierende Zielfunktion

minimiere $3x_1 + 5x_2$

da das erste Verfahren 3€ und das Zweite 5€ kostet.

Die Anwendung des ersten Verfahrens liefert uns 2 Einheiten S, die Anwendung des zweiten Verfahrens 1 Einheit S. Die Gesamtmenge S die produziert werden soll, muss der Lieferverpflichtung entsprechen, also mindestens 3 betragen. Somit können wir folgende Nebenbedingung an das schwere Öl stellen:

$$2x_1 + x_2 \geq 3$$

Analog verfahren wir für M und L. Somit ergeben sich die folgenden Nebenbedingungen:

$$\begin{array}{rcrcl} 2x_1 & + & x_2 & \geq & 3 \\ 2x_1 & + & 2x_2 & \geq & 5 \\ x_1 & + & 4x_2 & \geq & 4 \\ x_1 & & & \geq & 0 \\ & & x_2 & \geq & 0 \end{array}$$

Die letzten beiden Gleichungen sind notwendig, da wir ein Verfahren nicht eine negative Anzahl oft anwenden können.

(b) Die Umwandlung in Standardform ist recht einfach:

$$\textbf{maximiere } -3x_1 - 5x_2$$

unter den Nebenbedingungen

$$\begin{array}{rcrcl} -2x_1 & - & x_2 & \leq & -3 \\ -2x_1 & - & 2x_2 & \leq & -5 \\ -x_1 & - & 4x_2 & \leq & -4 \\ x_1 & & & \geq & 0 \\ & & x_2 & \geq & 0 \end{array}$$

(c) Siehe Vorlesungsfolien für eine graphische Präsentation.

Wir konstruieren uns zunächst das konvexe Polyeder aus den 5 Nebenbedingungen. Es fällt auf, dass das Polyeder nicht beschränkt ist. Das ist jedoch nicht weiter schlimm, da es in Richtung des Zielvektors c beschränkt ist. Wir beginnen nun bei der Ecke $(x_1, x_2) = (0, 3)$ mit dem Simplex–Algorithmus. Eine verbessernde Kante führt uns zum Punkt $(0.5, 2)$. Von hier aus gibt es eine weitere verbessernde Kante zum Punkt $(2, 0.5)$. Jede Kante von dieser Ecke aus wäre verschlechternd, also ist $(2, 0.5)$ unser Optimalwert[2].

Es fällt natürlich sofort auf, dass das Ergebnis unrealistisch ist. Man kann das Verfahren 2 nicht 0.5 mal anwenden. Würde man zusätzlich zu den Nebenbedingungen fordern, dass die Lösung ganzzahlig sein muss[3], so wird das Problem $\mathcal{NP}$-schwer und ist daher nicht mehr so einfach lösbar[4].

[2]Dieses lokale Optimum ist gleichzeitig das globale Optimum des Systems, da das System linear ist – bzw der Polyeder konvex.

[3]Was hier durchaus Sinn macht!

[4]Siehe dazu auch Info III.