Karlsruher Institut für Technologie
Institut für Theoretische Informatik

Prof. Dr. P. Sanders, Dr. S. Gog
L. Hübschle-Schneider, T. Maier

3. Zusatzblatt zu Theoretische Grundlagen der Informatik im WS 2015/16

http://algo2.iti.kit.edu/TGI2015.php
{sanders,gog,huebschle,t.maier}@kit.edu

Keine Abgabe, keine Korrektur!

Weitere Zusatzblätter: Weihnachtsblatt und Blatt 14

Aufgabe 1 (*NP-Vollständigkeitsreduktionen*)

Gegeben seien die folgenden Probleme:
ACYCLIC PARTITION: Gegeben ein *gerichteter* Graph $G = (V, E)$ und Zahlen $B, K, m \in \mathbb{N}$. Ist es möglich, V so in m disjunkte Teilmengen $V_1, \ldots, V_m$ zu zerlegen, sodass gilt:

- Die Größe jeder Partition ist höchstens B: $\forall_{i=1}^{m} |V_i| \leq B$
- Die Anzahl der Schnittkanten, d.h. solche die Partitionsgrenzen kreuzen, ist höchstens K: $|\{e = (u, v) \in E \mid\ u \in V_i,\ v \in V_j,\ i \neq j\} \leq K$
- Der "Partitionsgraph" G', der für jede der m Partitionen von V einen Knoten hat und eine gerichtete Kante zwischen zwei Knoten wenn es eine gerichtete Kante zwischen Knoten der entsprechenden Partitionen in G gibt, ist zyklenfrei?

EXACT 3COVER (X3C): Gegeben eine Menge X mit $|X| = 3q$ für $q \in \mathbb{N}$ und ein System X dreielementiger Teilmengen von X, d.h. $C \subseteq \mathcal{P}(Q)$ sodass $\forall c \in C : |c| = 3$. Gibt es eine Teilmenge $C' \subseteq C$ sodass jedes Element von X in genau einer Menge aus C' vorkommt, d.h. $\dot{\bigcup}_{c \in C'} c = X$?

Gegeben sei außerdem, dass EXACT 3COVER NP-vollständig ist (vergleiche: EXACT 3COVER zu EXACT COVER ist eine ähnliche Einschränkung wie 3SAT zu SAT; ohne Beweis). Zeigen Sie, dass ACYCLIC PARTITION ebenfalls NP-vollständig ist.
Hinweis: Überlegen Sie, wofür Sie Knoten einsetzen und in welche Richtung die Kanten zeigen müssen. Wie müssen Sie K und B wählen?

Musterlösung:
ACYCLIC PARTITION $\in$ NP: Gegeben einer potentiellen Lösung einer ACYCLIC PARTITION-Instanz können wir leicht überprüfen, ob diese die Bedingungen erfüllt. Die Balancebedingungen sind durch Summierung und Iteration über Knoten bzw Kanten schnell geprüft. Der Partitionsgraph lässt sich mit m Knoten für die m Partitionen initialisieren und durch Iteration über alle Kanten mit den notwendigen Kanten versehen. Seine Zyklenfreiheit kann mit einer Tiefensuche in Linearzeit verifiziert werden. Insgesamt können wir also in Polynomialzeit prüfen, ob ein gegebener Lösungskandidat das Problem tatsächlich löst, und daher ist ACYCLIC PARTITION in NP.

Reduktion von X3C: Gegeben eine X3C-Instanz (X, C) mit $|X| = 3q$. Wir erstellen für jedes Element von X einen Knoten, und für jede Menge aus C drei Knoten, die in einem Dreieck miteinander verbunden sind (Menge $\{a, b, c\}$ wird zu Kanten $(a, b), (b, c), (c, a)$. Wir nennen diese X- bzw. C-Knoten. Außerdem gibt es Kanten von den X-Knoten zu den C-Knoten, die die Vorkommen des Elements in den Mengen modellieren.

Wir setzen die maximale Partitionsgröße $K = 6$ und $B = |E| - 6q - 3(|C| - q)$. Diese Zahl ergibt sich daraus, dass wir q Partitionsmengen mit je 6 Kanten auswählen werden, und die anderen Mengen mit 3 C-Knoten werden ihre eigenen Partitionen haben.

Wenn es eine Überdeckung C' von X gibt, dann werden die zu den Mengen in der Überdeckung korrespondierenden C-Knoten mit den X-Knoten die sie überdecken gruppiert. Das ist unsere Partitionierung. Die übrigen Teilmengen von C bekommen entweder eigene Partitionen mit 3 Knoten oder können paarweise zu Partitionen mit 6 Knoten zusammengefasst werden (das ist egal). Wenn wir eine gültige Partitionierung haben, dann wissen wir:

- Keine Partition hat mehr als 6 Knoten (wegen Parameter K)
- Keine Partition kann die 3 C-Knoten, die eine Menge von C modellieren trennen, denn dann gäbe es einen Kreis im Partitionsgraphen.
- Jede Partition, die X-Knoten enthält, enthält genau 3 X-Knoten und 3 C-Knoten, die eine einzelne Teilmenge von C modellieren. Ansonsten würden nicht genug Schnittkanten aus dem Graph entfernt, denn der Parameter B wurde genau so gewählt, dass dies erzwungen wird. Beispielsweise würde eine Partition mit 6 X-Knoten keine potentiellen Schnittkanten aus dem Graphen entfernen, was es unmöglich machen würde, höchstens B Schnittkanten zu erhalten.

Mit dieser Konstruktion können wir aus der Lösung der EXACT 3COVER-Instanz einfach eine Lösung der ACYCLIC PARTITION-Instanz konstruieren, und andersherum aus einer Lösung der ACYCLIC PARTITION-Instanz eine Lösung der EXACT 3COVER-Instanz ablesen. Da die Reduktion polynomiell ist, folgt, dass ACYCLIC PARTITION NP-schwer ist.

Da wir gezeigt haben, dass ACYCLIC PARTITION in NP liegt und außerdem NP-schwer ist, ist es NP-vollständig. □

Aufgabe 2 (*NP-Vollständigkeit*)

Zeigen Sie: Wenn P=NP, dann sind genau alle nichttrivialen Sprachen aus P NP-vollständig. Nichttrivial bedeutet hier, dass die Sprache weder leer ist ($\emptyset$) noch alle Wörter enthält (Σ^*).

Musterlösung:
Offensichtlich sind $\emptyset$ und Σ^* in diesem Szenario nicht NP-vollständig: Eine Maschine die $\emptyset$ entscheidet akzeptiert kein Wort, und eine die Σ^* entscheidet lehnt keines ab. Es gibt also keine Möglichkeit, Ja-Instanzen anderer Probleme durch Instanzen von $\emptyset$ und Nein-Instanzen mittels Instanzen von Σ^* auszudrücken. Diese Probleme können also nicht NP-vollständig sein.

Alle anderen Probleme lassen sich beliebig polynomiell reduzieren, denn wir können das Problem bereits in der Reduktion lösen (ist ja in Polynomialzeit möglich!) und einfach eine triviale Ja- bzw Nein-Instanz des Zielproblems produzieren. Beispielsweise wäre eine Reduktion von SAT auf IS42 (für Eingabe $n \in \mathbb{N}$, gilt $n = 42$?) folgendermaßen möglich: Wenn die SAT-Instanz erfüllbar ist (prüfe dies während der Reduktion in polynomieller Zeit, denn P=NP), dann gib die IS42-Instanz 42 aus, andernfalls 1337. Insofern handelt es sich fast schon um eine Trickfrage!

Aufgabe 3 (*Wissensfragen*)

Gelten die folgenden Aussagen? Begründen Sie kurz!

a) Wenn NP=PSPACE, dann muss P≠NP gelten

b) Wenn ein Optimalwertproblem von einer nichtdeterministischen Turingmaschine in polynomieller Zeit gelöst werden kann, ist das zugehörige Entscheidungsproblem in NP

c) Nichtdeterministische Turingmaschinen können kontextsensitive Sprachen in Polynomialzeit entscheiden.

Musterlösung:

a) Unfug. Es ist bekannt: $P \subseteq NP \subseteq PSPACE$. Es ist möglich, dass keine dieser Inklusionen echt ist (d.h. $P = PSPACE$), das ist nur ziemlich unwahrscheinlich.

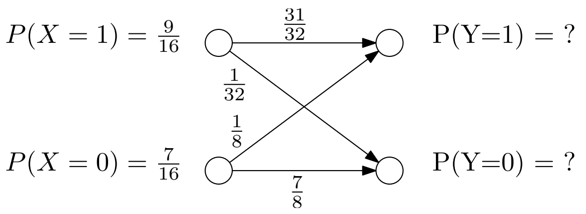
b) Ja, es muss zusätzlich nur noch geprüft werden ob der Optimalwert die zusätzlichen Bedingungen des Entscheidungsproblems (typischerweise von der Form "Optimalwert ist mindestens X" oder "Optimalwert ist höchstens X") erfült.

c) Unbekannt (vermutlich nicht). Das Chomsky-1-Wortproblem ist PSPACE-vollständig, und die Menge der kontextsensitiven Sprachen ist echte Teilmenge von PSPACE.

Aufgabe 4 (*Informationstheorie (Alphabetcodierung + Verrauschter Kanal)*)

Gegeben sei das folgende Alphabet mit Buchstabenwahrscheinlichkeiten, und der binäre Kanal:

A	B	C	D	E	F
0.3	0.15	0.1	0.2	0.15	0.1



a) Um die Buchstaben über den binären Kanal zu versenden, müssen diese zunächst binär codiert werden. Berechnen Sie hierzu einen entsprechenden Huffman-Baum/Code.

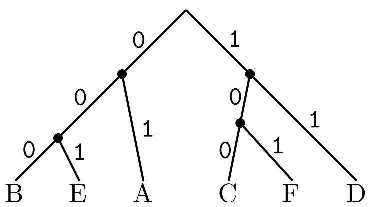
b) Berechnen Sie nun die Entropie der Quelle ($H(X)$) und der Senke ($H(Y)$). Berechnen Sie außerdem den übertragenen Informationsgehalt ($I(X;Y)$). Hierzu benötigen Sie die bedingten Entropien $H(X|Y)$ und $H(Y|X)$, als auch die kombinierten Wahrscheinlichkeiten $P(X = x, Y = y)$.

c) Wie lässt sich der in (a) berechnete Huffman-Code variieren, sodass auf dem gegebenen Kanal die Fehlerwahrscheinlichkeit eines gesendeten F minimal wird? (Der resultierende Code soll weiterhin ein Huffman-Code für die gegebenen Buchstabenwahrscheinlichkeiten sein)

d) Was ist die Fehlerwahrscheinlichkeit eines gesendeten C in dem variierten Huffman-Code?

Musterlösung:

a)



Korrekturhinweis: Im Allgemeinen ist der Huffman-Baum nicht eindeutig, in diesem Fall ist er bis auf Isomorphie eindeutig. Dementsprechend sind in diesem Beispiel auch die Codelängen jedes Zeichens eindeutig.

b)

$$
\begin{aligned}
P(Y=1) &= \frac{7}{128} + \frac{279}{512} = \frac{307}{512} \\
P(Y=0) &= \frac{49}{128} + \frac{9}{512} = \frac{205}{512} \\
H(X) &= -\sum_x P(X=x)\log P(X=x) \\
&= -\frac{9}{16}\log\frac{9}{16} - \frac{7}{16}\log\frac{7}{16} \\
&= -\frac{9}{16}(\log 9 - 4) - \frac{7}{16}(\log 7 - 4) = 4 - \frac{9}{16}\log 9 - \frac{7}{16}\log 7 \approx 0.989 \\
H(Y) &= -\sum_y P(Y=y)\log P(Y=y) \\
&= -\frac{307}{512}\log\frac{307}{512} - \frac{205}{512}\log\frac{205}{512} = 9 - \frac{307}{512}\log 307 - \frac{205}{512}\log 205 \approx 0.971 \\
H(Y|X) &= -\sum_x\sum_y P(X=x,Y=y)\log P(Y=y|X=x) \\
&= -\frac{279}{512}\log\frac{31}{32} - \frac{9}{512}\log\frac{1}{32} - \frac{7}{128}\log\frac{1}{8} - \frac{49}{128}\log\frac{7}{8} \\
&= \frac{279}{512}\cdot 5 + \frac{9}{512}\cdot 5 + \frac{7}{128}\cdot 3 + \frac{49}{128}\cdot 3 - \frac{279}{512}\log 31 - \frac{49}{128}\log 7 \\
&= \frac{33}{8} - \frac{279}{512}\log 31 - \frac{49}{128}\log 7 \approx 0.351 \\
P(X=1|Y=1) &= \frac{279}{512}/\frac{307}{512} = \frac{279}{307} \\
P(X=0|Y=1) &= \frac{28}{512}/\frac{307}{512} = \frac{28}{307} \\
P(X=1|Y=0) &= \frac{9}{512}/\frac{205}{512} = \frac{9}{205} \\
P(X=0|Y=0) &= \frac{196}{512}/\frac{205}{512} = \frac{196}{205} \\
H(X|Y) &= -\sum_x\sum_y P(X=x,Y=y)\log P(X=x|Y=y) \\
&= -\frac{279}{512}\log\frac{279}{307} - \frac{7}{128}\log\frac{28}{307} - \frac{9}{512}\log\frac{9}{205} - \frac{49}{128}\log\frac{196}{205} \approx 0.368 \\
I(X;Y) &= H(X) - H(X|Y) = H(Y) - H(Y|X) \approx 0.621
\end{aligned}
$$

c) Wie in Aufgabenteil (a) beschrieben ist die Codelänge in diesem Beispiel eindeutig. Daher lässt sich die Fehlerwahrscheinlichkeit nicht durch einen kürzeren Code verringern. Da die Fehlerwahrscheinlichkeit abhängig ist von dem gesendeten Zeichen (Fehlerawahrscheinlichkeit bei $X=1 \Rightarrow \frac{1}{32}$ bei $X=0 \Rightarrow \frac{1}{8}$) ist es optimal möglichst nur `1`en zu senden. Deshalb variieren wir die in (a) berechneten Codes, indem wir die Kantenbeschriftung an einer Stelle verändern. Dadurch ergibt sich $Code(F) =$ `111` $Code(C) =$ `110` $Code(D) =$ `10`, die anderen Codes bleiben unverändert.

d) In jedem Huffman-Code mit $C(F) = 111$ hat ein gesendetes C immer den Code `110` (da es auf der untersten Ebene mit F verknüpft ist). Die Fehlerwahrscheinlichkeit ist deshalb $1 - \frac{31}{32}\cdot\frac{31}{32}\cdot\frac{7}{8} \approx 0.179 \approx 18\%$