Rev. 1, 15.06.2018
TU Ilmenau, Fakultät für Informatik und Automatisierung
FG Komplexitätstheorie und Effiziente Algorithmen
Univ.-Prof. Dr. M. Dietzfelbinger, M. Sc. P. Schlag
http://www.tu-ilmenau.de/iti/lehre/lehre-ss-2018/ra/

TECHNISCHE UNIVERSITÄT
ILMENAU

Randomisierte Algorithmen SS 2018 – Übung 6

Besprechung: Montag, 25. Juni 2018

Hinweis: Für das erfolgreiche Vorrechnen einer mit „*“ gekennzeichneten Aufgabe wird ein Bonuspunkt vergeben, es gibt maximal zwei Bonuspunkte pro Studierendem im Semester. Bitte geben Sie Ihren Lösungsvorschlag bis zum Vortag der Übung (Sonntag), 13:00 Uhr, per E-Mail an philipp.schlag@tu-ilmenau.de oder bis Freitag, 15:00 Uhr, direkt in meinem Büro (Z 1048) ab.

Aufgabe 1 (Count-Min-Sketch) *

Sei $(x_1, x_2, \ldots, x_n)$ eine Folge von n Objekten aus $A := \{1, 2, \ldots, N\}$ mit $N \geq 1$ und sei $c(x)$ die absolute Häufigkeit von $x \in A$ in der gegebenen Folge. Die Häufigkeiten $c(x)$ sollen mit $o(N)$ Zählern geschätzt werden. Dazu benutzt Algorithmus 1 ein Array $\mathtt{A}[1..d, 1..w]$ aus $d \cdot w$ Zählern und d Hashfunktionen $h_1, h_2 \ldots, h_d : A \to \{1, \ldots, w\}$. Die Hashfunktionen erfüllen $\mathbf{Pr}(h_i(x) = h_i(y)) \leq \frac{1}{w}$ für alle $1 \leq i \leq d$ und Schlüssel $x \neq y$ und sind unabhängig voneinander.

Algorithmus 1 : Count-Min-Sketch

Eingabe : Folge $(x_1, x_2, \ldots, x_n)$ über A

Ausgabe : Array $\mathtt{A}[1..d, 1..w]$ mit $A[j,k] = |\{1 \leq i \leq n \mid h_j(x_i) = k\}|$ für $1 \leq j \leq d$, $1 \leq k \leq w$

erzeuge Array $\mathtt{A}[1..d, 1..w]$ mit Eintrag 0 in jeder Arrayzelle;
for $i = 1$ **to** n **do**
  for $j = 1$ **to** d **do** $\mathtt{A}[j, h_j(x_i)] \leftarrow \mathtt{A}[j, h_j(x_i)] + 1$;
end

Für $x \in A$ ergibt sich dann der Schätzwert

$$\hat{c}(x) := \min\{\mathtt{A}[j, h_j(x)] \mid 1 \leq j \leq d\}.$$

Wir zeigen nun, dass $\hat{c}(x)$ bei geeigneter Wahl von d und w mit mindestens einer gewissen Wahrscheinlichkeit in einem kleinen Toleranzband um $c(x)$ liegt.

(a) Zeigen Sie: Für alle $x \in A$ gilt $\hat{c}(x) \geq c(x)$.

(b) Seien $x \in A$ und $j \in \{1, 2, \ldots, d\}$ beliebig, fest. Wir definieren die Zufallsvariable

$$Z := \sum_{y \in A \setminus \{x\}} [h_j(y) = h_j(x)] \cdot c(y).$$

Zeigen Sie **(i)** $\mathtt{A}[j, h_j(x)] = c(x) + Z$ und **(ii)** $\mathbf{E}(Z) \leq \frac{n}{w}$.

Weiter auf nächster Seite!

(c) Zeigen Sie: Für jedes $x \in A$ und $\varepsilon > 0$ gilt

$$\mathbf{Pr}(\hat{c}(x) \geq c(x) + \varepsilon n) = \mathbf{Pr}(Z \geq \varepsilon n)^d \leq (\varepsilon w)^{-d}.$$

(d) Zeigen Sie, dass sich aus der Wahl $d = \lceil \log_b \frac{1}{\delta} \rceil$, für ein $0 < \delta < 1$ und $b > 1$, sowie $w = \lceil \frac{b}{\varepsilon} \rceil$ der Parameter d und w Folgendes ergibt:

$$\mathbf{Pr}(c(x) \leq \hat{c}(x) < c(x) + \varepsilon n) \geq 1 - \delta.$$

(e) Seien $0 < \delta < 1$ und $\varepsilon > 0$ gegeben. Die Parameterwahl in (d) liefert uns damit einen Schätzwert $\hat{c}(x)$, der mit einer Wahrscheinlichkeit von höchstens δ um mindesten εn von der tatsächlichen Häufigkeit $c(x)$ abweicht. Der Speicherverbrauch $d \cdot w$ von Algorithmus 1 ist dabei proportional zu $(\log_b \frac{1}{\delta}) \cdot \frac{b}{\varepsilon}$. Wählen Sie $b > 1$ derart, dass dieser möglichst klein wird!

Aufgabe 2 (Es funkt zwischen Alice und Bob) *

Alice und Bob möchten über einen Funkkanal kommunizieren. Bob sendet das Signal $b \in \{-1, 1\}$. Der Sendepegel -1 repräsentiert dabei ein Null-Bit und der Pegel 1 repräsentiert ein Eins-Bit. Neben Bob gibt es $n \geq 2$ weitere Sender, die in gleicher Weise kodierte Signale $b_1, \dots, b_n$ senden. Alice empfängt

$$a := b + \sum_{i=1}^{n} w_i \cdot b_i \text{ mit } 0 < w_i \leq 1 \text{ für } 1 \leq i \leq n \text{ und } W := \sum_{i=1}^{n} w_i > 1,$$

wobei die Gewichte w_i Dämpfungen der anderen Signale (wegen Entfernung, Hindernissen, etc.) modellieren. Alice interpretiert $a < 0$ als ein Null-Bit und $a \geq 0$ als ein Eins-Bit.
Im Folgenden nehmen wir an, dass Bob $b = 1$ (also ein Eins-Bit) sendet und die Signale $b_1, \dots, b_n$ uniform verteilt und unabhängig voneinander sind. Als zugrundeliegenden Wahrscheinlichkeitsraum verwenden wir daher $\Omega = \{-1, 1\}^n$ mit $p_\omega = 2^{-n}$. Die Störsignale $b_1, \dots, b_n$ sind dann binäre Zufallsvariablen mit Wertebereich $\{-1, 1\}$, das empfangene Signal a ist eine reellwertige Zufallsvariable mit Wertebereich $[-n+1, n+1]$. Wir wollen eine obere Schranke für die Wahrscheinlichkeit $\mathbf{Pr}(a < 0)$, dass Alice das Signal fälschlicherweise als Null-Bit interpretiert, bestimmen.

(a) Zeigen Sie, dass gilt:

$$a \leq 0 \iff \sum_{i=1}^{n} \frac{w_i}{W} \cdot \frac{b_i + 1}{2} \leq \frac{1}{2} \cdot \left(1 - \frac{1}{W}\right).$$

(b) Zeigen Sie für die Fehlerwahrscheinlichkeit:

$$\mathbf{Pr}(a < 0) \leq \left[e^{\frac{1}{2W}} \left(1 - \frac{1}{W}\right)^{\frac{1}{2}\left(1 - \frac{1}{W}\right)} \right]^{-1} =: f(W).$$

Hinweis: Nutzen Sie (a) und wenden Sie die Hoeffding-Ungleichung (Korollar 2.6.3 (2)) an.

(c) Zeigen Sie, dass $f(W)$, $1 < W < \infty$, streng monoton wachsend ist und bestimmen Sie $\lim_{W \searrow 1} f(W)$ sowie $\lim_{W \to \infty} f(W)$.