Kryptologie

Die mittlere Zeichenkoinzidenz zweier stochastischer Sprachen

```
a7Hzq .#5r<
kÜ\as TâÆK$
ûj(Ö2 ñw%h:
Úk{4R f~`z8
¤˜Æ+Ô „&¢Dø
```

Für stochastische Sprachen $L, M \subseteq \Sigma^*$ mit Buchstabenhäufigkeiten q_s bzw. p_s für $s \in \Sigma$ wird die mittlere Zeichenkoinzidenz von Texten der Länge r betrachtet:

$$\kappa_{LM}^{(r)} := \frac{1}{\#L_r} \cdot \frac{1}{\#M_r} \cdot \sum_{a \in L_r} \sum_{b \in M_r} \kappa(a,b)$$

Satz. *Die mittlere Zeichenkoinzidenz der stochastischen Sprachen L und M ist asymptotisch gleich*

$$\lim_{r \to \infty} \kappa_{LM}^{(r)} = \sum_{s \in \Sigma} p_s q_s$$

Der *Beweis* folgt unten.

Deutung

Die Zeichenkoinzidenz genügend langer gleichlanger Texte $a \in L$ und $b \in M$ ist ungefähr

$$\kappa(a,b) \approx \Sigma_{s \in \Sigma}\, p_s q_s.$$

Das stimmt überein mit der intuitiven Vorstellung, wie wahrscheinlich das Auftreten von Koinzidenzen (Zwillingspaaren) ist.

Spezialfälle

1.) Sei $L = \Sigma^*$ mit den Buchstabenhäufigkeiten $q_s = 1/n$, und M habe die Buchstabenhäufigkeiten p_s. Dann ist

$$\kappa_{M\Sigma^*} = \Sigma_{s \in \Sigma}\, p_s/n = 1/n.$$

Das deutet man so:

$$\kappa(\text{»sinnvoller Text«}, \text{»zufälliger Text«}) \approx 1/n.$$

2.) Sei $L = M$. Dann erhält man die Formel

$$\kappa_{MM} = \Sigma_{s \in \Sigma}\, p_s^2.$$

Das deutet man so:

$$\kappa(\text{»sinnvoller Text«}, \text{»sinnvoller Text«}) \approx \Sigma_{s\in\Sigma}\, p_s^2.$$

3.) Sei $L = M_{(q)} = \{a_{(q)} \mid a \in M\} \subseteq \Sigma^*$; L besteht also aus den um q Stellen zyklisch verschobenen Texten. Dann ist mit M auch L stochastisch, und zwar mit den gleichen Buchstabenhäufigkeiten. Also ist

$$\kappa_{LM} = \Sigma_{s\in\Sigma}\, p_s^2.$$

Für die Texte $a \in M$ bilden die Paare $(a,a_{(q)})$ allerdings keine »repräsentative« Stichprobe aus $L\times M$. Nimmt man aber an, dass $a_{(q)}$ »unabhängig« von a ist - was bei natürlichen Sprachen schon bei $q \geq 2$ empirisch möglich ist - so erhält man die Näherungsformel

$$\kappa_q(a) \approx \Sigma_{s\in\Sigma}\, p_s^2.$$

Ein Hilfssatz

Hilfssatz. *Sei M eine stochastische Sprache. Dann gilt für die mittlere Abweichung für alle Buchstaben* $s \in \Sigma$:

$$\frac{1}{r}\cdot\sum_{j=0}^{r-1}\left(\mu_{sj}^{(r)} - p_s\right) \to 0 \quad \text{für } r \to \infty.$$

Beweis. Sei $\varepsilon > 0$ gegeben und r so groß, dass

a) $r \geq 4 \cdot \#J/\varepsilon$,

b) $|\mu_{sj}^{(r)} - p_s| < \varepsilon/2$ für alle $j \in [0 \ldots r]\text{-}J$.

Für $j \in J$ ist sicher $|\mu_{sj}^{(r)} - p_s| \leq |\mu_{sj}^{(r)}| + |p_s| \leq 2$. Also folgt:

$$\frac{1}{r}\cdot\sum_{j=0}^{r-1}|\mu_{sj} - p_s| < \frac{1}{r}\cdot 2\cdot\#J + \frac{r-\#J}{r}\cdot\frac{\varepsilon}{2} \leq \frac{\varepsilon}{2}+\frac{\varepsilon}{2} = \varepsilon.$$

♦

Bemerkung.

$$\mu_s^{(r)} = \frac{1}{r}\cdot\sum_{j=0}^{r-1}\mu_{sj}^{(r)} = \frac{1}{r}\cdot\frac{1}{\#M_r}\cdot\sum_{a\in M_r}\delta_{sa_j}$$

ist die mittlere Häufigkeit von s in Texten der Länge r. Dafür gilt also:

Korollar. $\lim_{r\to\infty} \mu_s^{(r)} = p_s$.

Der Beweis des Satzes

$$\kappa_{LM}^{(r)} = \frac{1}{\#L_r \cdot \#M_r} \cdot \sum_{a\in L_r} \sum_{b\in M_r} \left(\frac{1}{r} \cdot \sum_{j=0}^{r-1} \sum_{s\in\Sigma} \delta_{sa_j} \cdot \delta_{sb_j} \right)$$

$$= \sum_{s\in\Sigma} \frac{1}{r} \cdot \sum_{j=0}^{r-1} \left[\frac{1}{\#L_r} \cdot \sum_{a\in L_r} \delta_{sa_j} \right] \cdot \left[\frac{1}{\#M_r} \sum_{b\in M_r} \delta_{sb_j} \right]$$

$$= \sum_{s\in\Sigma} \frac{1}{r} \cdot \sum_{j=0}^{r-1} \left[q_s + \varepsilon_{sj} \right] \cdot \left[p_s + \eta_{sj} \right]$$

$$= \sum_{s\in\Sigma} \left[p_s q_s + \frac{p_s}{r} \cdot \sum_{j=0}^{r-1} \varepsilon_{sj} + \frac{q_s}{r} \cdot \sum_{j=0}^{r-1} \eta_{sj} + \frac{1}{r} \cdot \sum_{j=0}^{r-1} \varepsilon_{sj}\, \eta_{sj} \right]$$

Der zweite und dritte Summand konvergieren nach dem Hilfssatz gegen 0, der vierte konvergiert ebenfalls gegen 0, da $|\varepsilon_{sj}\eta_{sj}| \le 1$. Also konvergiert die Summe gegen $\Sigma_{s\in\Sigma}\, p_s q_s$. ♦

Autor: Klaus Pommerening, 5. März 2000; letzte Änderung: 6. März 2000.
E-Mail an Pommerening »AT« imbei.uni-mainz.de.