Kryptologie

Der Koinzidenzindex einer stochastischen Sprache

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

Definition

Für eine stochastische Sprache $M \subseteq \Sigma^*$ mit Buchstabenhäufigkeiten p_s heißt

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

der **Koinzidenzindex** von M (nach FRIEDMAN).

Deutung: Für unabhängige Texte $a, b \in M$ gleicher Länge ist

$$\kappa(a,b) \approx \kappa_M$$

Beispiele. 1.) $\kappa_{\Sigma^*} = 1/n$. Im Spezialfall $n = 26$ ist $\kappa_{\Sigma^*} \approx 0.0385$.

2.) Aus den bekannten Häufigkeitstabellen folgen empirisch die bereits angekündigten Werte

- für M = »Deutsch«: $\kappa_M \approx 0.0762$,
- für M = »Englisch«: $\kappa_M \approx 0.0661$.

Eigenschaften

Da $\Sigma_{s\in\Sigma}\, p_s = 1$, gilt $1/n \le \kappa_M \le 1$, und zwar

a. $\kappa_M = 1/n \Leftrightarrow$ alle $p_s = 1/n$,
b. $\kappa_M = 1 \Leftrightarrow$ ein $p_s = 1$, alle übrigen = 0.

Das folgt aus

a. $1 = (\Sigma_{s\in\Sigma}\, p_s\cdot 1)^2 \le \Sigma_{s\in\Sigma}\, p_s^{\,2} \cdot \Sigma_{s\in\Sigma}\, 1 = n \cdot \Sigma_{s\in\Sigma}\, p_s^{\,2}$

   mit Gleichheit $\Leftrightarrow (p_s)_{s\in\Sigma} = c\cdot(1)_{s\in\Sigma}$ (als Vektor) $\Leftrightarrow$ alle p_s gleich $\Leftrightarrow$ alle $p_s = 1/n$.

b. $\Sigma_{s\in\Sigma}\, p_s^{\,2} \le \Sigma_{s\in\Sigma}\, p_s = 1$, da $0 \le p_s \le 1$, also $p_s^{\,2} \le p_s$,

   mit Gleichheit $\Leftrightarrow$ alle $p_s^{\,2} = p_s \Leftrightarrow$ alle $p_s = 0$ oder 1.

Anwendungen

1.) Bei polyalphabetischer Substitution ist die Wiederverwendung des gleichen Schlüssels mit hoher

Wahrscheinlichkeit erkennbar (egal, ob periodisch oder nicht):

Für *verschiedene* polyalphabetische Substitutionen f, g ist zu erwarten, dass

$$\kappa(f(a),g(b)) \approx 1/n \quad \text{für } a, b \in \Sigma^r.$$

Jetzt ist auch geklärt, dass bei *gleicher* Substitution

$$\kappa(f(a),f(b)) = \kappa(a,b) \approx \kappa_M \quad \text{für } a, b \in M.$$

2.) Sei c ein Geheimtext aus einer polyalphabetischen Verschlüsselung eines Textes $a \in M$ der Periode l. Welche Werte sind für $\kappa_q(c)$ zu erwarten?

	$c = c_0$	...	c_{q-1}	||	c_q	...	c_{r-1}
	$c^{(q)} = c_{r-q}$	...	c_{r-1}	||	c_0	...	c_{r-q-1}
erwartete Koinzidenzen:	$q\cdot\kappa_M$, falls		$l\|r-q$,	||	$(r-q)\cdot\kappa_M$, falls		$l\|q$,
	$q\cdot\kappa_{\Sigma^*}$ sonst,			||	$(r-q)\cdot\kappa_{\Sigma^*}$ sonst.		

Daraus ergeben sich folgende erwartete Werte für den Autokoinzidenzindex:

1. Fall, $l|r$:

$$\kappa_q(c) \approx \begin{cases} \dfrac{q\cdot\kappa_M + (r-q)\cdot\kappa_M}{r} = \kappa_M, & \text{falls } l|q, \\[2ex] \dfrac{q\cdot\kappa_{\Sigma^\times} + (r-q)\cdot\kappa_{\Sigma^\times}}{r} = \kappa_{\Sigma^\times} & \text{sonst.} \end{cases}$$

2. Fall, l kein Teiler von r:

$$\kappa_q(c) \approx \begin{cases} \dfrac{q\cdot\kappa_{\Sigma^\times} + (r-q)\cdot\kappa_M}{r} & \text{falls } l|q, \\[2ex] \dfrac{q\cdot\kappa_M + (r-q)\cdot\kappa_{\Sigma^\times}}{r} & \text{falls } l|r-q, \\[2ex] \kappa_{\Sigma^\times} & \text{sonst.} \end{cases}$$

Insbesondere gilt für $q << r$:

$$\kappa_q(c) \approx \begin{cases} \kappa_M, & \text{wenn } l|q, \\ \kappa_{\Sigma^*} & \text{sonst.} \end{cases}$$

Damit ist auch das im Beispiel beobachtete Autokoinzidenzspektrum erklärt, und die typischen Autokoinzidenzspektren sehen so aus.

Autor: Klaus Pommerening, 6. März 2000; letzte Änderung: 30. November 2004.
E-Mail an `Pommerening »AT« imbei.uni-mainz.de`.