TECHNISCHE UNIVERSITÄT DRESDEN

Professur Datenschutz und Datensicherheit

Fakultät Informatik, Institut für Systemarchitektur, Professur Datenschutz und Datensicherheit

barkhauseninstitut.org

Komplexpraktikum — Friedman-Test

dud.inf.tu-dresden.de

Stefan Köpsell (stefan.koepsell@tu-dresden.de)

Koinzidenzindex

- Wahrscheinlichkeit, dass zwei zufällig aus Text gewählte Buchstaben gleich sind
- Symbol: I

$$I = \sum_{i=1}^{n} p_i^2$$

n: Anzahl Zeichen im Alphabet

p_i: Auftrittswahrscheinlichkeit für i. Zeichen aus Alphabet

Koinzidenzindex

- Wahrscheinlichkeit, dass zwei zufällig aus Text gewählte Buchstaben gleich sind
- Symbol: I

$$I = \sum_{i=1}^{n} p_i^2$$

- Koinzidenzindex ist minimal, wenn alle Zeichen gleichwahrscheinlich vorkommen:

$$I_{min} = \sum_{i=1}^{n} \left(\frac{1}{n}\right)^2 = n\frac{1}{n^2} = \frac{1}{n}$$

Koinzidenzindex

- Koinzidenzindex für deutschsprachigen Text:

$$I_{Deutsch} = 0{,}0762$$

- Koinzidenzindex zufälliger Text aus 26 Buchstaben

$$I_{26} = 0{,}038462 < I_{Deutsch}$$

- Hinweis:
  - Koinzidenzindex unabhängig von „Benennung" der Zeichen
    - nur Auftrittswahrscheinlichkeit zählt

➔ **MM-Substitution ändert Koinzidenzindex nicht!**

Bestimmung der Schlüssellänge mit Hilfe des Koinzidenzindex

⌘ Idee: Schlüssellänge l gruppiert Schlüsseltext in l viele MM-Substitutionen

➔ bei richtig bestimmter Schlüssellänge sollte pro Gruppe der Koinzidenzindex etwa $I_{Deutsch}$ sein, andernfalls eher I_{min}

Schlüsselzeichen	k_1	k_2	k_3
Beispiel	**K**	**E**	**Y**
Schlüssel-text-zeichen	c_1	c_2	c_3
	c_4	c_5	c_6
	c_7	c_8	c_9

$I_{Deutsch}$

⌘ naive Lösung: Schlüssellänge l durchprobieren

⌘ besser: Schlüssellänge l analytisch bestimmen

Bestimmung der Schlüssellänge mit Hilfe des Koinzidenzindex

Schlüsselzeichen	k_1	k_2	k_3
Beispiel	**K**	**E**	**Y**
Schlüssel-text-zeichen	c_1	c_2	c_3
	c_4	c_5	c_6
	c_7	c_8	c_9

- Vorbetrachtungen:
  - Anzahl Zeichenpaare P_{Ges} gesamter Schlüsseltext der Länge N:
    - $$P_{Ges} = \frac{N(N-1)}{2}$$
  - Anzahl Zeichenpaare P_{Spalte} pro Spalte:
    - $$P_{Spalte} = \frac{\frac{N}{l}\left(\frac{N}{l}-1\right)}{2} = \frac{\frac{N}{l}\left(\frac{N}{l}-\frac{l}{l}\right)}{2} = \frac{N(N-l)}{2l}$$

Bestimmung der Schlüssellänge mit Hilfe des Koinzidenzindex

- Vorbetrachtungen:
  - Anzahl Zeichenpaare P_{Ges} gesamter Schlüsseltext der Länge N:
    - $$P_{Ges} = \frac{N(N-1)}{2}$$
  - Anzahl Zeichenpaare P_{Spalte} pro Spalte:
    - $$P_{Spalte} = \frac{\frac{N}{l}\left(\frac{N}{l}-1\right)}{2} = \frac{\frac{N}{l}\left(\frac{N}{l}-\frac{l}{l}\right)}{2} = \frac{N(N-l)}{2l}$$
  - Anzahl Zeichenpaare P_V verschiedene Spalte:
    - $$P_V = \frac{N^2(l-1)}{2l}$$

Bestimmung der Schlüssellänge mit Hilfe des Koinzidenzindex

- Wahrscheinlichkeit für Paar mit gleichen Zeichen (=Koinzidenzindex I):
  - Merke:
    - pro Spalte: Wahrscheinlichkeit=I_{Deutsch}
    - ansonsten: Wahrscheinlichkeit=I_{min}

$$I = \frac{P_{Spalte} \cdot I_{Deutsch} + P_V \cdot I_{min}}{P_{Ges}} = \frac{\frac{N(N-l)}{2l} \cdot I_{Deutsch} + \frac{N^2(l-1)}{2l} \cdot I_{min}}{\frac{N(N-1)}{2}}$$

- Umstellen nach l:

$$I = \frac{\frac{N(N-l)}{2l} \cdot I_{Deutsch} + \frac{N^2(l-1)}{2l} \cdot I_{min}}{\frac{N(N-1)}{2}} = \frac{\frac{N(N-l)}{l} \cdot I_{Deutsch} + \frac{N^2(l-1)}{l} \cdot I_{min}}{N(N-1)}$$

$$I \cdot N \cdot (N-1) = \frac{N(N-l)}{l} \cdot I_{Deutsch} + \frac{N^2(l-1)}{l} \cdot I_{min}$$

$$I \cdot N \cdot (N-1) \cdot l = N(N-l) \cdot I_{Deutsch} + N^2(l-1) \cdot I_{min}$$

$$I \cdot \quad (N-1) \cdot l = \quad (N-l) \cdot I_{Deutsch} + N \ (l-1) \cdot I_{min}$$

$$I \cdot \quad (N-1) \cdot l = N \cdot I_{Deutsch} - l \cdot I_{Deutsch} + Nl \cdot I_{min} - N \cdot I_{min}$$

Bestimmung der Schlüssellänge mit Hilfe des Koinzidenzindex

- Wahrscheinlichkeit für Paar mit gleichen Zeichen (=Koinzidenzindex I):
  - Merke:
    - pro Spalte: Wahrscheinlichkeit=I_{Deutsch}
    - ansonsten: Wahrscheinlichkeit=I_{min}
  - $I \cdot (N-1) \cdot l = N \cdot I_{Deutsch} - l \cdot I_{Deutsch} + Nl \cdot I_{min} - N \cdot I_{min}$
  - $I \cdot (N-1) \cdot l + l \cdot I_{Deutsch} - Nl \cdot I_{min} = N \cdot I_{Deutsch} - N \cdot I_{min}$
  - $I \cdot (N-1) \cdot l + l \cdot I_{Deutsch} - Nl \cdot I_{min} = N(I_{Deutsch} - I_{min})$
  - $(I \cdot (N-1) + I_{Deutsch} - N \cdot I_{min})l = N(I_{Deutsch} - I_{min})$
  - $$l = \frac{N(I_{Deutsch} - I_{min})}{I \cdot (N-1) + I_{Deutsch} - N \cdot I_{min}}$$

Bestimmung der Schlüssellänge mit Hilfe des Koinzidenzindex

- Wahrscheinlichkeit für Paar mit gleichen Zeichen (=Koinzidenzindex I):
  - Merke:
    - für gesamten Schlüsseltext: $I = \sum_{i=1}^{n} \left(\frac{h_i}{N}\right)^2$
      - mit h_i: Häufigkeit des i. Zeichen im Schlüsseltext
    - pro Spalte: Wahrscheinlichkeit=I_{Deutsch}
    - ansonsten: Wahrscheinlichkeit=I_{min}

$$l = \frac{N(I_{Deutsch} - I_{min})}{I \cdot (N-1) + I_{Deutsch} - N \cdot I_{min}}$$