

Übung zur Vorlesung "Computerlinguistik II / Sprachtechnologie"

Sommersemester 2018, Prof. Dr. Udo Hahn, Tobias Kolditz
Übungsblatt 7 vom 20.06.2018
Abgabe bis 27.06.2018 per E-Mail (PDF-Datei) an
tobias.kolditz@uni-jena.de

Aufgabe 1 : Inter-Annotator Agreement

7 pt

Das *inter-annotator agreement* beschreibt die Ähnlichkeit mehrerer unterschiedlicher Annotationen auf dem gleichen Text. Ein solches Agreement-Maß für kategoriale Urteile von zwei Annotatoren ist *Cohen's Kappa*. Sei U eine $u \times u$ Urteilmatrix, die insgesamt N Urteile enthält, so berechnet sich Cohen's Kappa κ wie folgt:

$$(1) \quad \kappa = \frac{p_a - p_e}{1 - p_e}$$

Wobei p_a die beobachtete Übereinstimmung zwischen beiden Annotatoren darstellt:

$$(2) \quad p_a = \frac{\sum_{i=1}^u U_{ii}}{N}$$

Und p_e , die zufällig erwartete Übereinstimmung, wie folgt berechnet wird:

$$(3) \quad p_e = \frac{1}{N^2} \sum_{i=1}^u \left(\sum_{j=1}^u U_{ji} \cdot \sum_{j=1}^u U_{ij} \right)$$

Betrachten Sie folgende Statistik für zwei Annotationen von NP-Chunks im IO Format:

A_1/A_2	I	O
I	40	10
O	20	30

a) Berechnung von Cohen's Kappa

3 pt

Berechnen Sie Cohen's Kappa für die obige Statistik. Geben Sie Ihren Lösungsweg an!

b) Ergebnisdiskussion

2 pt

Weist das Ergebnis auf eine hohe oder niedrige Übereinstimmung der Tag-Vergabe hin? Deckt sich das Ergebnis mit dem, was Sie anhand der Tabelle erwartet hätten? Begründen Sie Ihre Antwort.

c) Probleme mit Cohen's Kappa

2 pt

Berechnen Sie Cohen's Kappa für die Urteilmatrizen in Abbildung 1, adaptiert aus Byrt et al. (1993).¹ Vergleichen Sie die Ergebnisse für (a) mit (b) sowie (c) mit (d). Wie beeinflussen die verschiedenen Datenverteilungen den Kappa-Wert?

Aufgabe 2 : Brat Stand-off Format

3 pt

Informieren Sie sich auf <http://brat.nlplab.org/standoff.html> über das von Brat genutzte Standoff-Format und wandeln Sie folgendes XML-Beispiel (im Inline-Format) in dieses Format um.

```
<entity id="1" type="ORG">BWM</entity> produziert den <entity id="2" type="MISC">BMW i3</entity> in <entity id="3" type="LOC">Leipzig</entity>.
```

¹In dieser ‚abstrakten Annotationsaufgabe‘ musste jeder Annotator jeweils 100 Ja-/Nein-Urteile abgeben.

A_1/A_2	Yes	No	Total
Yes	40	20	60
No	20	20	40
Total	60	40	100

(a) Gleiche Randverteilung bei beiden Annotatoren (60% Yes)

A_1/A_2	Yes	No	Total
Yes	40	35	75
No	5	20	25
Total	45	55	100

(b) Unterschiedliche Randverteilungen (45% vs 75% Yes)

A_1/A_2	Yes	No	Total
Yes	40	10	50
No	10	40	50
Total	50	50	100

(c) Gesamtverteilung über Kategorien ist uniform

A_1/A_2	Yes	No	Total
Yes	70	10	80
No	10	10	20
Total	80	20	100

(d) Prävalenz einer Klasse (Yes) deutlich höher

Abbildung 1: Urteilsmatrizen für Aufgabe 1c

Literatur

Byrt, T., Bishop, J., & Carlin, J. B. (1993). Bias, prevalence and kappa. *Journal of Clinical Epidemiology*, 46(5), 423-429.