

Exposé für eine Bachelorarbeit: Untersuchung der Effizienz eines Assistenzsystems für Textannotation

Robert Greinacher (33xxxx), Informatik Bsc. (Fakultät IV)

Einleitung / Problemstellung

Zur maschinellen Begriffserkennung von Personen- oder Firmennamen in Texten muss ein Computer zunächst auf die Erkennung der gesuchten Worte / Begriffe / Wortarten und -gruppen trainiert werden. Hierfür werden Trainingsdatensätze benötigt, anhand derer ein repräsentierendes Modell erzeugt wird, um eigenständig Text zu klassifizieren. Trainingsdaten sind annotierte Texte deren Worte (alle oder einige derer) je nach Anwendung mit Attributen versehen sind, die einer Markierung der gesuchten Worte gleichkommt. Anhand dieser Attribute kann ein Algorithmus lernen, Struktur- und Wortzusammenhänge zu erkennen.

Um diese Trainingsdatensätze zu generieren annotieren üblicherweise Menschen die, je nach Aufgabe, relevanten Worte und Satzstrukturen händisch. Annotationen mit bestehenden Werkzeugen sind sehr zeit- und dadurch kostenintensiv und bedürfen bei den Annotatoren viel Konzentration und Durchhaltevermögen.

Das Annotationsinterface

Um Annotationen durchzuführen wurde ein simples Annotationsinterface entwickelt. Dieses präsentiert einzelne Absätze des Textkorpus als ein zu annotierendes Dokument (im Folgenden Annotationsdokument genannt). Dabei kann ein Annotationsdokument mehrere Annotationsstellen beinhalten. Ist ein Annotationsdokument fertig bearbeitet, wird es gespeichert und das nächste geladen.

Unterstützung durch ein Assistenzsystem

Ein Assistenzsystem soll Annotatoren durch Vorannotation der Annotationsdokumente unterstützen. Eine perfekte Assistenz liefert demnach ein vollständig und korrekt annotiertes Annotationsdokument. Ohne die Assistenz ist ein zu bearbeitendes Annotationsdokument vollständig unannotiert.

Das Assistenzsystem wird in der Realität fehlerbehaftete Vorannotationen liefern. Generell, ob von Nutzern oder von einem Assistenzsystem, können sich Annotationen in sechs verschiedenen Aspekten (Antwortklassen) unterscheiden – wobei nur einer dieser einer korrekten Annotation entspricht (s. Anhang).

Die geplante Studie untersucht den Vorteil des Einsatzes dieser Assistenz. Konkret werden die Fehler einer Annotation ohne Assistenz mit den Fehlern unter Einsatz des Assistenzsystems verglichen und die entstandene Fehlervarianz untersucht.

Das Versuchsdesign

Es gibt vier Stufen des Assistenzsystems; 10% korrekter Assistenz (und entsprechend 90% fehlerhaften Annotationen), mit 50% korrekter Assistenz mit 90% korrekter Assistenz und ohne Assistenz. Letzteres bedeutet, dass die VPs ein komplett unannotiertes Dokument erhalten und alle Annotationen selbst finden und einstellen (= Kategorie wählen, Länge und Ende definieren). Diese Stufe stellt die „Baseline“ der Untersuchung dar.

In einer der Stufen *mit Assistenz* sind anteilig (10% / 50% / 90%) alle Annotationsstellen eines Dokuments durch das Assistenz-System korrekt vorannotiert. Die übrigen Annotationsstellen sind ebenfalls vorannotiert, enthalten jedoch Fehler (alle Möglichkeiten, eine Annotationsstelle fehlerhaft zu machen, werden gleich häufig ausgenutzt; somit sind alle verschiedenen Fehler gleichverteilt¹). In diesen Assistenz-Stufen müssen die VPs also die Annotation des Assistenzsystems überprüfen und ggf. korrigieren.

Somit ergibt sich ein vierstufiges, einfaktorielles Versuchsdesign mit folgenden Variablen:

UV

Stufe des Assistenzsystems

Alle VPs durchlaufen zwei der vier Assistenz-Stufen : Jeweils genau eine der drei unterschiedlich starken Ausprägungen der Assistenz, sowie der Baseline-Bedingung:

- Baseline bzw. eine der drei Stufen (10%, 50%, 90%) der Stärke der Assistenz als Within-Subjects-Design (in Blöcken von je 5 Dokumenten mit und ohne Assistenz)
- die Zugehörigkeit zu einer der drei Assistenz-Stufen mit Vorannotation (Between-Subjects), somit ist die Korrektheit der Vorannotationen pro VP konstant.

Vorannotation des Assistenzsystems (verschachtelte Stufen; betrifft nur Blöcke mit Assistenz)

Within-Subjects-Design über alle Antwortklassen;
randomisiert und gleich verteilt, jedoch unter Berücksichtigung der Assistenz-Stufe. Die fehlerhaften Annotationen des Assistenzsystems werden also über alle möglichen Fehler verteilt und gespeichert, um später eine Aussage über das Korrekturverhalten der VPs treffen zu können.

AV

Fehler der Nutzer an jeder einzelnen Annotationsstelle

Operationalisierung: Die von den VPs annotierten Annotationsdokumente werden mit einem annotierten Referenz-Annotationsdokument (einer perfekten Lösung, 100% korrekte Annotationen) verglichen. Durch den Vergleich aller Annotationsstellen mit der Referenz ergibt sich eine Antwortklasse pro Annotationsstelle.

¹ Es ist davon auszugehen, dass die Auftretenshäufigkeit der fehlerbehafteten Antwortklassen in der Realität nicht gleich verteilt sein wird, weder von einem Assistenzsystem, noch von den VPs (nachzuweisen durch Analyse der Annotationen der Kontrollgruppe / Stufe ohne Vorannotation). Hierzu liegen jedoch keine genauen Zahlen vor, da diese von einer konkreten Implementierung des Assistenzsystems abhängen. Dieser Gedanke ist als Hypothese formuliert.

% absolut gefundener Annotationen

Operationalisierung: Die von den VPs annotierten Annotationsdokumente werden mit einem annotierten Referenz-Annotationsdokument verglichen. Durch die Differenz der Annotationen ergibt sich die Anzahl der ausgelassenen (übersehenen) Annotationen durch die VPs.

Außerdem werden die von den VPs annotierten Annotationsdokumente mit der Version vor der Annotation durch die VPs verglichen. Durch diese Differenz ergibt sich die Anzahl der von den VPs getätigten (im Falle der Stufe mit AS) oder veränderten (im Falle der Stufe ohne AS) Annotationen.

Anzahl Fehler pro Fehlerklasse

Operationalisierung: Für jede Annotationsstelle wird die Fehlerklasse aufsummiert; es ergibt sich die Anzahl der Fehler pro Fehlerklasse. (Hat Abhängigkeit zur *UV mit / ohne Assistenz*.)

% korrekte Annotationen

Operationalisierung: Bei dem Vergleich mit der Referenzannotation wird nur diejenige Annotation der VPs als gültig gewertet, die vollständig mit der Referenz übereinstimmt. Jede Abweichung zur Referenz ist demnach eine fehlerhafte Annotation. Abschließend ergibt sich ein Verhältnis durch n / m mit n = Anzahl korrekter Annotationen und m = Anzahl aller Annotationsstellen

durchschnittlich Annotationszeit

Operationalisierung: Die Gesamtzeit t der Annotation von n Annotationsdokumenten wird erhoben. Diese n Annotationsdokument enthalten insgesamt m Annotationsstellen. Somit bildet $t' = m / t$ die durchschnittliche Zeit pro Annotation.

Hypothesen

Das 10% korrekte Assistenzsystem unterstützt die Annotation:

- Mit AS werden Dokumente schneller annotiert als ohne AS.
- Mit AS werden mehr Annotationen richtig annotiert als ohne AS.
- Mit AS werden weniger Annotationsstellen übersehen als ohne AS.

Das 50% korrekte Assistenzsystem unterstützt die Annotation:

- Mit AS werden Dokumente schneller annotiert als ohne AS.
- Mit AS werden mehr Annotationen richtig annotiert als ohne AS.
- Mit AS werden weniger Annotationsstellen übersehen als ohne AS.

Das 90% korrekte Assistenzsystem unterstützt die Annotation:

- Mit AS werden Dokumente schneller annotiert als ohne AS.
- Mit AS werden mehr Annotationen richtig annotiert als ohne AS.
- Mit AS werden weniger Annotationsstellen übersehen als ohne AS.

Manche fehlerbehafteten Antwortklassen lassen sich besser korrigieren als andere

Beispiel: Eine nicht vorhandene Annotation könnte leichter übersehen werden, als eine unvollständige vorhandene.

Operationalisierung: Für jede Antwortklasse wird eine Liste erstellt, wie häufig eine Vorannotation dieser Klasse zu welcher anderen Antwortklasse korrigiert wurde. Darauf lässt sich pro Antwortklasse der Vorannotation ablesen, in wie vielen Fällen diese Vorannotation richtig korrigiert werden konnte.

Ablauf je VP

- Einführung & Interface-Demonstration
- Training des Umgangs mit dem Interface (bis qualitatives Kriterium erfüllt; Beispiel: 9 / 10 Annotationsstellen werden korrekt annotiert). Soll Umgang mit dem Interface erleichtern und einen Lerneffekt über die Zeit der Benutzung während der Studie minimieren. Es werden unannotierte, mit dem Versuch vergleichbare Texte (bzgl. Inhalt / Komplexität und Länge) von den VPs annotiert.
- Versuch; 50 Annotationsdokumente in Blöcken zu je 5 Dokumenten. Wechselweise 5 Dokumente mit Assistenzsystem und 5 Dokumente ohne AS. Insgesamt werden 25 Dokumente mit Assistenz und 25 ohne Assistenz präsentiert.

Anhang

Sechs Antwortklassen der Annotationen

Eine Annotation ist ein Tupel von Worten, zusammen mit einer Kategorie (Person oder Firma) an einer Annotationsstelle. Eine fehlerhafte Annotation kann daher ein Tupel falscher Länge oder mit falschem Beginn sein, oder der falschen Kategorie angehören. Somit ist jede Annotation Element einer der folgenden Klassen:

#	Kategorie	Spanne	Beschreibung	Beispiele
0	✓	✓	Alles korrekt	Die Geschäftsführerin Susanne Spong tanzt gern.
1	✓	✗		Die Geschäftsführerin Susanne Spong tanzt gern. Die Geschäftsführerin Susanne Spong tanzt gern. Die Geschäftsführerin Susanne Spong tanzt gern.
2	✗	✓		Die Geschäftsführerin Susanne Spong tanzt gern.
3	✗	✗		Die Geschäftsführerin Susanne Spong tanzt gern.
4	✗	✗	Unnötige Annotation	Die Geschäftsführerin Susanne Spong tanzt gern.
5	–	–	Keine Annotation vorhanden	Die Geschäftsführerin Susanne Spong tanzt gern.

Unnötige Annotation, ID 4: Im Unterschied zu ID 3 ist die Annotation kein Teil einer Annotationsstelle (es "gehört" keine Annotation an diese Stelle).

Diese sechs Klassen sind exklusiv. Ein Assistenzsystem, dessen Annotationen bezüglich der Antwortklassen gleich verteilt wären, würde zu 16,666% (= Häufigkeit der ID 0 Klasse) korrekt arbeiten.