

Textual Entailment

23. Januar 2008

Prinzipien des Corpus Calculus

Lukas Bulwahn
bulwahn@in.tum.de

Überblick

- Was ist Textual Entailment
 - Definition und Beispiele, prakt. Anwendung
 - The RTE Challenge
- Lösungsansätze für Textual Entailment
 - Word-of-Bag Modelle
 - syntaktische Modelle
 - semantische Modelle
- Zusammenfassung

Definition von Textual Entailment

- Definition von Textual Entailment (Sprachliches Folgern):

Eine Hypothese H folgt aus Text T , falls normalerweise ein Mensch, der T liest, folgert, dass H höchst wahrscheinlich wahr ist.

Definition von Textual Entailment

- Ein einfaches Beispiel für TE:

A 30-year-old man has been killed in a shark attack at a surfing beach near Perth in West Australia where he was surfing with four other people.

A 30-year-old man was killed in a shark attack while surfing.

Definition von Textual Entailment

- Ein einfaches Beispiel für TE: **(TRUE)**

A 30-year-old man has been killed in a shark attack at a surfing beach near Perth in West Australia where he was surfing with four other people.

A 30-year-old man was killed in a shark attack while surfing.

Definition von Textual Entailment

- Ein weiteres einfaches Beispiel:

Japan's voter turnout was just over 56 percent
for the Upper House elections.

Less than half of the eligible Japanese voters
participated in the vote.

Definition von Textual Entailment

- Ein weiteres einfaches Beispiel: **(FALSE)**

Japan's voter turnout was just over 56 percent
for the Upper House elections.

Less than half of the eligible Japanese voters
participated in the vote.

Definition von Textual Entailment

- Ein einfaches Beispiel für TE:

The two suspects belong to the 30th Street gang, which became embroiled in one of the most notorious recent crimes in Mexico: a shootout at the Cuadalajara airport in May, 1993, that killed Cardinal Juan Jesus Posadas Ocampo and six others.

Cardinal Juan Jesus Posadas Ocampo died in 1993.

Definition von Textual Entailment

- Ein einfaches Beispiel für TE: **(TRUE)**

The two suspects belong to the 30th Street gang, which became embroiled in one of the most notorious recent crimes in Mexico: a shootout at the Cuadalajara airport in May, 1993, that killed Cardinal Juan Jesus Posadas Ocampo and six others.

Cardinal Juan Jesus Posadas Ocampo died in 1993.

Für das Publikum

- **Wahr oder Falsch?**

The first settlements on the site of Jakarta were established at the mouth of the Ciliwung, perhaps as early as the 5th century AD.

Für das Publikum

- **Wahr oder Falsch?**

The first settlements on the site of Jakarta were established at the mouth of the Ciliwung, perhaps as early as the 5th century AD.

The first settlements on the site of Jakarta were established as early as the 5th century AD.

- Und nun ist das Publikum gefragt:
- Nochmals zur Erinnerung die Definition:

Eine Hypothese H folgt aus Text T, falls normalerweise ein Mensch, der T liest, folgert, dass H höchst wahrscheinlich wahr ist.

Probleme der Definition

- Definition ist relativ informell
- daher wurden auch weitere Richtlinien festgelegt:
 - Alle Teile von H sollten aus T zu erschließen sein. Falls nicht, ist die Folgerung falsch.
 - Falls die Folgerung sehr wahrscheinlich ist (aber nicht absolut sicher), ist sie wahr.
 - Man kann einfaches Allgemeinwissen verwenden. (z.B. an CEO is an employee of the company.)

Anwendung in der Praxis

- Question-Answering-System
(“die semantische Suchmaschine”)
 - Anfragen in natürlicher Sprache
“What is the height of the Eiffel Tower?”
 - Extraktion der passenden Antwort:
Aus Textsammlung muss “The Eiffel Tower is 300 meters tall” als Antwort gefunden werden.
- Auswertung maschineller Übersetzung
 - Die automatische Übersetzung eines Textes folgt aus dem Original und vice versa.

The RTE Challenge

- Schreibe ein Programm, das T und H als Eingabe nimmt, und ausgibt, ob T aus H folgt
- Sicherlich eine der schwierigsten Aufgaben der Computerlinguistik und Künstlichen Intelligenz
- Auswertung der Programme anhand von 800 (400 wahren und 400 falschen) kurzen Beispielpaaren (und 567 Paaren zum Üben)

The RTE Challenge

- Eine universelle, anwendungsunabhängige Aufgabenstellung
- schafft eine Vergleichsmöglichkeit für Ansätze aus verschiedenen Bereichen
 - z. B. QA und MT-Systeme, Information Extraction
- schafft eine Vergleichsmöglichkeit für verschiedene semantische Theorien der Linguistik

Verschiedene Ansätze bei RTE

- Wir werden nun drei verschiedene Ansätze betrachten:
 - Analyse auf lexikalischer Ebene
 - Analyse auf syntaktischer Ebene
 - Analyse auf semantischer Ebene

Der WEM-Algorithmus

- Der einfachste und aktuell "beste" Ansatz:
 - Algorithmus:
 - 1. Schritt: Lese die Texte T und H.
 - 2. Schritt: **W**irf **E**ine **M**ünze
 - 3. Schritt: Bei Kopf antworte "wahr"; bei Zahl antworte "falsch"
 - 1. Schritt ist natürlich optional!
 - Erfolgsquote: **50%**

Using BLEU (Perez and Alfonseca, 2005)

- BLEU zur Auswertung von maschinellen Übersetzung gegen menschl. Übersetzung
- misst die Anzahl der übereinstimmenden N-gramme der zwei Texte
- BLEU ergibt ein Ergebnis zwischen 0 und 1
- Empirisch gewählte Schranke trennt zwischen wahren und falschen Paaren
- Erfolgsquote: **51 - 53%**

Beispiele für BLEU/BoW

- Beispiel: **(TRUE)**

The Clark County medical examiner's office said **the man** who was killed was **33** years old.

The Clark County medical examiner's office put **the dead man's** age at **33**.

Beispiele für BLEU/BoW

- Beispiel: **(FALSE)**

Militant groups in **Iraq have** waged a kidnapping campaign aimed at driving out U.S. supporting companies and troops.

About 70 foreigners **have** been abducted by militants fighting the U.S.-led occupation and reconstruction efforts in **Iraq**.

Gegenbeispiel für BLEU/BoW

- Beispiel: **(FALSE)**

The economy created 228,000 new jobs after a disappointing 112,000 in June.

The economy created 228,000 jobs after dissapointing the 112,000 of June.

Bag of Words Modelle

- Betrachte alle Paare von vorkommenden Wörtern in T und H
- Finde Paare mit maximaler Ähnlichkeit
- Gewichte die Paare nach Relevanz (inverse document frequency)
- Summiere zu einem Ergebnis und entscheide
- Ähnlichkeit anhand von: gemeinsamen Vorkommen in Dokumenten, WordNet-Abstand
- Erfolgsquote: **55 - 60%**

Probabilistic Textual Entailment (Glickman)

- Wir wollen feststellen, ob die Wahrscheinlichkeit, dass H wahr ist unter der Bedingung, dass T wahr ist, "ausreichend groß" ist.
- Dazu:
$$Pr(Tr_h = 1 | t) = \prod_{u \in h} \max_{v \in t} Pr(u | v)$$

Probabilistic Textual Entailment (Glickman)

- Wir wollen feststellen, ob die Wahrscheinlichkeit, dass H wahr ist unter der Bedingung, dass T wahr ist, "ausreichend groß" ist.

- Dazu:
$$Pr(Tr_h = 1 | t) = \prod_{u \in h} \max_{v \in t} Pr(u | v)$$

- Wir approximieren mit einem Korpus:

$$Pr(u | v) = \frac{\text{Anz. der Dokumente mit } u \text{ und } v}{\text{Anz. der Dokumente mit } v}$$

Beispiel

Japan's voter turnout was just 56 percent for the Upper House elections.

Less than half of the eligible Japanese voters participated in the vote.

Beispiel

Japan's voter turnout was just 56 percent for the Upper House elections.

Less than half of the eligible Japanese voters participated in the vote.

Beispiel

Japan's voter turnout was just 56 percent for the Upper House elections.

Less than half of the eligible Japanese voters participated in the vote.

Beispiel

Japan's voter turnout was just 56 percent for the Upper House elections.

Less than half of the eligible Japanese voters participated in the vote.

Beispiel

Japan's voter turnout was just 56 percent for the Upper House elections.

Less than half of the eligible Japanese voters participated in the vote.

Fazit: Manche Paare sind sinnvoll –
die meisten jedoch einfach statischer Unsinn!

Syntaktische Modelle

- Transformiere T und H in syntaktische Struktur
- Berechne Kosten der syntaktischen Transformation um von T zu H zu gelangen
- Verschiedene Ansätze: Tree-Edit-Distance, Teilgraph-Erkennung, Analyse von Dependency Trees
- Erfolgsquote: **55 - 59%**

Der klassische Ansatz

- Übersetzen von T und H in eine formale semantische Form
- logisches Schließen in einem Kalkül mit klar definierten Herleitungsregeln (z.B. FOL)
- Verwende dazu FOL Solver (z. B. Otter)
- Entscheidend ist hier, welches weitere Wissen verwendet wird, um Beweise zu finden.
- Beispiele: Montague-Grammatik, DRT usw.

Ausblick: Schließen in der Logik

- Prädikatenlogik hat eine klar definierte Semantik und Schlußregeln.
- Daher ist es einfach zu zeigen, dass z. B.

$$\frac{\exists x. \forall y. P(x, y)}{\forall y. \exists x. P(x, y)}$$

- Für die Prädikatenlogik existieren effiziente Verfahren, um lange Beweise innerhalb von Sekunden automatisch zu lösen.

Semantik bei RTE (Bos & Markert)

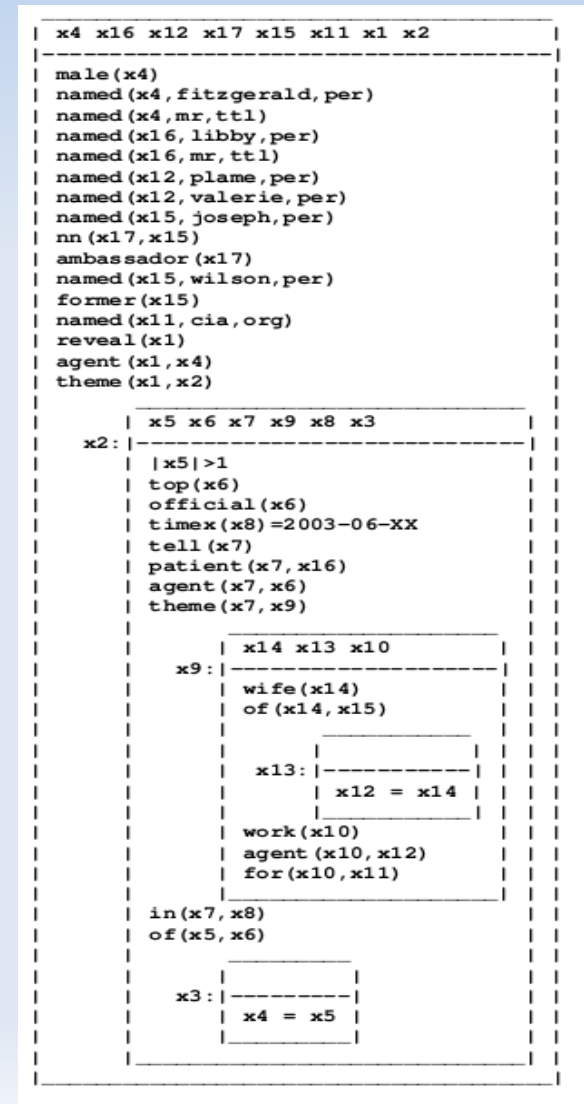
- Text wird in DRT transformiert
- Hintergrundwissen: linguistische Transformationen (wie Passiv), WordNet, CIA factbook (geograph. Wissen)
- Verwenden von OTTER zum Suchen von Beweisen (nur 29 Beweise bei 800 Beispielen)
- Verwenden von Paradox oder Mace um Modelle zu erstellen

Transformation in DRT

Mr. Fitzgerald revealed he was one of several top officials who told Mr. Libby in June 2003 that Valerie Plame, wife of the former ambassador Joseph Wilson, worked for the CIA.

Valerie Plame is married to Joseph Wilson.

x3 x1 x2
named(x3,plame,per)
named(x3,valerie,per)
marry(x1)
patient(x1,x3)
named(x2,wilson,per)
named(x2,joseph,per)
to(x1,x2)



Zusammenfassung von RTE

- Momentaner Stand der Technik:
 - viele verschiedene Ansätze zur Erkennung von Schlußfolgerungen
 - viele absolut sprachunabhängig
 - manche absolut sinnlos, aber alle schneiden gleichermaßen schlecht ab.
 - Aktuelle Teilerfolge aufgrund von Verwendung größerer Ressourcen (wie WordNetX, oder Sammlungen von Weltwissen)

Literatur

- Proceedings to the RTE Challenge Workshops
- Oren Glickman. Applied Textual Entailment.

Vielen Dank für die Aufmerksamkeit. **noch Fragen?**