

HS Textual Entailment

Überblick

Peter Kolb

Wintersemester 2008/2009

Variabilität der Sprache

- gleiche Bedeutung kann durch unterschiedliche Texte ausgedrückt oder aus unterschiedlichen Texten inferiert (gefolgert) werden:

Regierungschefin hält Rede zur Entwicklung der Wirtschaft

Bundeskanzlerin thematisiert konjunkturelle Situation

Merkel spricht über wirtschaftliche Aussichten

- → Paraphrasieproblem für NLP-Anwendungen, wird für jede Anwendung einzeln angegangen

„Recognizing Textual Entailment“

- allgemeine, abstrakte Aufgabenstellung für das Folgern aus Texten
- Framework aus Task-Definition und Evaluierungsmethode
- PASCAL Recognizing Textual Entailment (RTE) Challenge
- Erster RTE-Workshop 2005
- Ziel: Entwicklung eigenständiger RTE-Module, die wie Tagger oder Parser in NLP-Anwendungen integrierbar sind

RTE: Aufgabenstellung

- Text T, Hypothese H (Satz, Phrase, Wort)
- angewandtes Folgern aus Text: $T \Rightarrow H$, wenn Leser aus T schließt, dass H höchstwahrscheinlich wahr ist
- „höchstwahrscheinlich“: Abkehr von strikter logischer Inferenz
- operationale Definition
- setzt menschliches Sprachvermögen und Weltwissen voraus

Definitionen für „Folgerung“

- klassisches logisches Folgern: Hypothese H folgt aus Text T, wenn H in jeder möglichen Welt wahr ist, in der T wahr ist.
- wahrscheinlichkeitstheoretische Definition:
 $T \Rightarrow H$, wenn $P(H \text{ ist wahr} | T) > P(H \text{ ist wahr})$
gegebenes T erhöht Wahrscheinlichkeit, dass H wahr ist
- $T + \text{ sprachl. Wissen} + \text{ Weltwissen} \Rightarrow H$

Vielseitige Relation

- Mittels Folgerungsrelation kann Paraphrasie definiert werden:

$T \Leftrightarrow H$ (Paraphrase, Äquivalenz) wenn gilt:
 $T \Rightarrow H$ und $H \Rightarrow T$

- Folgerung als Relation zwischen Texten definiert \rightarrow unabhängig von bestimmter semantischer Interpretation
- z.B.: Lesartendisambiguierung: Problem Lesarteninventar. Stattdessen prüfen, ob sich im Kontext ein anderes Wort folgern lässt:

Vielseitige Relation (2)

IKEA announced a new comfort chair

chair => furniture

MIT announced a new CS chair position

chair ≠> furniture

Zuweisung einer best. Lesart unnötig

- Begriffshierarchien: *Hund* Hyperonym von *Pudel* wenn *Alle Hunde bellen => Alle Pudel bellen*

RTE in NLP-Anwendungen

- QA: *Was produziert Peugeot?*

Chretien besuchte Peugeots frisch renovierte Autofabrik => Peugeot stellt Autos her

- IR: Trefferdokument => Anfrage

- IE: H: Firma stellt Produkt her

*Der Software-Hersteller SAS veröffentlichte...
=> H*

Milupa, Produzent von Babynahrung, ... => H

RTE in NLP-Anwendungen (2)

- SUM: Satz in Zusammenfassung, der aus anderen Sätzen in Zusammenfass. gefolgert werden kann, ist redundant
- CAT: Suche nach passender Übersetzung in Translation Memory
- MT: Übersetzung soll mit Vorgabe aus Goldstandard semantisch äquivalent sein

PASCAL RTE Challenge

- RTE-1 2005: Genauigkeit 52%-70%
- 567 T-H-Paare Training, 800 Test (RTE-1), bei 50% der Paare folgt H aus T, 50% nicht.
- RTE-2 2006: Genauigkeit 53%-75%
- RTE-3 2007 auch YES/NO/UNKNOWN
- Daten aus Anwendungsszenarios:
 - IR: Sätze aus News als Anfrage an Suchmaschine geschickt (H), in Ergebnissen T ausgewählt.
 - CD: News-Cluster
 - sowie RC, QA, IE, MT, PP

Beispieldaten (1)

- *The Arabic-language television network Al-Jazeera reports it has received a statement and a videotape from militants.*
- *Al-Jazeera is an Arabic-language television network.*
- (folgt.)

Beispieldaten (2)

- *Security authorities have declared a state of maximum emergency in Guatemala, which is located directly in the path of the hurricane.*
- *There is a state of maximum emergency in Guatemala because of the hurricane.*
- (folgt.)

Beispieldaten (3)

- *With \$549 million in cash as of June 30, Google can easily afford to make amends.*
- *Some 30 million shares have been assigned to the company's workers.*
- (folgt nicht.)

Beispieldaten (4)

- *Even though the rug was ripped out from under Lollapalooza this week, Jane's Addiction frontman Perry Farrell is optimistic that the eclectic music festival he co-founded in 1991 will resurface and it might be sooner than you think.*
- *Perry Farrel hopes to found another music festival.*
- (folgt nicht.)

Ein erster Ansatz

- Monz & de Rijke (2001)
- Voraussetzung für Folgerung: verarbeitbare Repräsentation des Textinhalts
- klassischerweise Prädikatenlogik erster Stufe (PL1)
- aber: Umwandlung nat. Sprache in PL1 erfordert volles Parsing inkl. Disambiguierung
- nicht robust, zu teuer, keine Implementation

Ein erster Ansatz (2)

- selbst wenn Umwandlung in PL1 gelingt:
- Folgerungsproblem in PL1 unentscheidbar
- Folgerung in PL1 entweder wahr oder falsch, keine Ausnahmen oder Wahrscheinlichkeiten
→ ungeeignet für Repräsentation vager nat. Sprache
- zwischen T und H in Daten gilt fast nie logisch strikte Folgerung
- lieber eine nur wahrscheinlich richtige Antwort als gar keine

Ein erster Ansatz (3)

- einfachere Repräsentation: nur partielles Parsing, Unterspezifikation zur Vermeidung vollst. Disambiguierung
- noch einfacher: Bag-of-Words-Repräsentation
- Monz & de Rijke messen Übereinstimmung (Ähnlichkeit) zwischen Wörtern in T und H
- falls Ähnlichkeit zwischen T und H über Schwellwert, gilt $T \Rightarrow H$.
- Genauigkeit auf RTE-1-Evaluierungsdaten 55%

Ein erster Ansatz (4)

- stellen Vorteile ihres Systems heraus:
 - ✓ vollständig implementiert & effizient
 - ✓ domänenunabhängig
 - ✓ Ausgabe mit Verlässlichkeitswert statt ja/nein
- in weiteren RTE-Challenges stoßen flache Systeme aber an 60%-Grenze
- erst wissensintensivere, aufwändigere Systeme erzielen höhere Genauigkeiten