Textual Entailment

Alina Petrova

EMCL TUD, HLT FBK

February 22, 2012

Introduction

Textual Entailment (TE):

- What is it?
  a notion from classical logic is applied to natural language using NLP technologies
- Which techniques can be applied?
  relevant features for detecting TE via machine learning
- What is done by the community?
  RTE Challenge

Introduction

Textual Entailment (TE):

- What is it?
  a notion from classical logic is applied to natural language using NLP technologies
- Which techniques can be applied?
  relevant features for detecting TE via machine learning
- What is done by the community?
  RTE Challenge

Fondazione Bruno Kessler, Human Language Technology group
RTE-7 Challenge participation

Natural Language Processing Nowadays

Definition

NLP is an interdisciplinary field which seeks to enable computer to process, understand and generate natural language.

Natural Language Processing Nowadays

Definition

NLP is an interdisciplinary field which seeks to enable computer to process, understand and generate natural language.

Modern NLP consists of multiple subareas which can be defined by the tasks they aim to solve.

- Machine Translation
- Information Retrieval
- Question Answering
- Word Sense Disambiguation
- ...
- **Recognizing Textual Entailment**

Textual Entailment

Intuition: Recognizing Textual Entailment is a generic task that captures major semantic inference between pieces of text.

Definition

Given two text fragments, *Text (T)* and *Hypothesis (H)*:
T entails H iff the meaning of H can be inferred from the meaning of T by human reading.

Textual Entailment

Intuition: Recognizing Textual Entailment is a generic task that captures major semantic inference between pieces of text.

Definition

Given two text fragments, *Text (T)* and *Hypothesis (H)*:
T entails H iff the meaning of H can be inferred from the meaning of T by human reading.

Notes:

- why "human reading"?
- what is a "text fragment"?

Example:
T: If you help the needy, God will reward you.
H: Giving money to a poor man has good consequences.

TE: How-To

2 opposite approaches:

Using formal sematics:

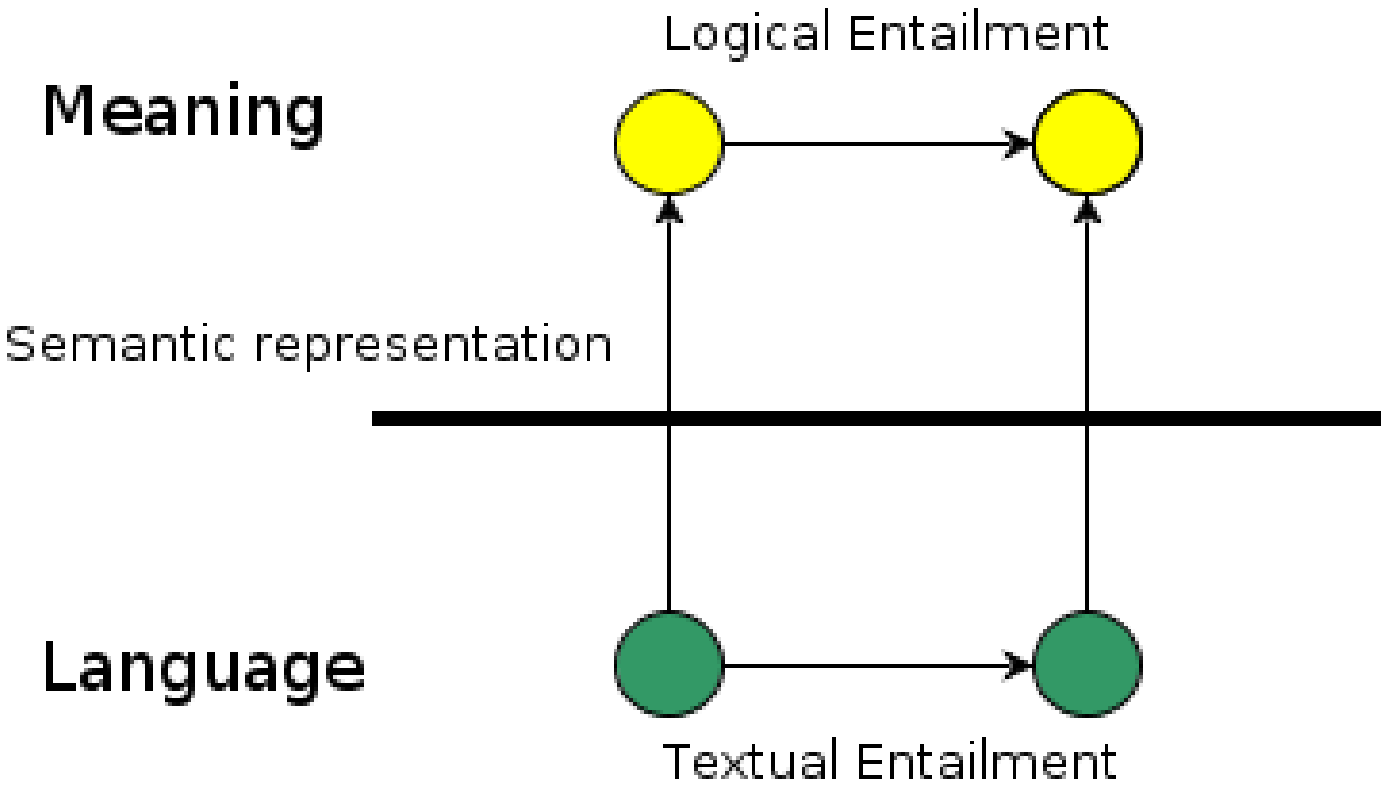
- translation of natural language fragments into some logical systems
- classical approach which brings together logic, language and psychology
- successful for narrow domains, but not working on comprehensive data!
- few training data

Using surface structure:

- counterintuitive, but proved to be fruitful.

Why? A wide range of entailments follow general patterns that arise from surface (lexical and syntactic) considerations.

TE: How-To cont'd



Surface approach

Main feature is **lexical similarity**.

- naive word overlap
- n-grams (= sequences of neighboring words) overlap

Ex: A student Computational Logic workshop took place in Vienna. ⇒ Workshop took place in Vienna.

- normalized forms

working = work, brought = bring

- paraphrasing (different lexical forms with similar meaning)

Ex: A student workshop was organised in the capital of Austria. ⇒ A student workshop took place in Vienna.

Surface Approach - cont'd

The entailment holds iff the word overlap reaches a certain threshold.
It is set via supervised learning.

Surface Approach - cont'd

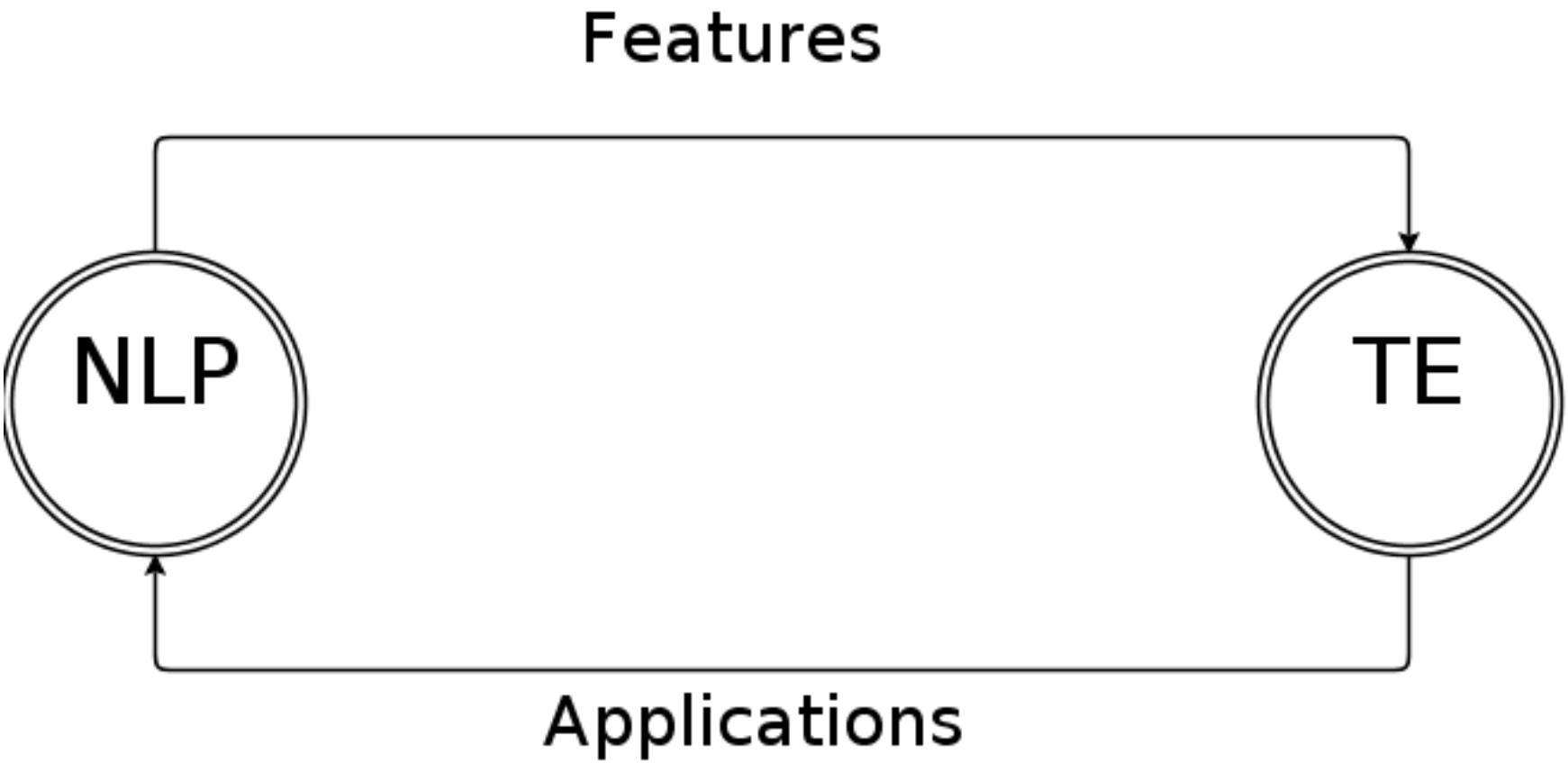
The entailment holds iff the word overlap reaches a certain threshold.
It is set via supervised learning.

Statistics on F-measure (2010 data):

- best performance - 48.01%
- average performance - 33.77%
- **up to 40% using only lexical matching**

But this seems to be a limit for lexical matching.

NLP vs. Textual Entailment



NLP contribution to TE

Using extra features from other areas of NLP improve lexical match results:

- Semantic Roles
- Named Entity Recognition
- lexical knowledge bases (VerbOcean, WordNet)
- coreference
- syntactic parsing

etc.

Applications

Textual entailment recognition is used in several NLP tasks:

- *Question Answering*
- Information Extraction
- Information Retrieval
- Text Summarization

and many more.

Applications

Textual entailment recognition is used in several NLP tasks:

- *Question Answering*
- Information Extraction
- Information Retrieval
- Text Summarization

and many more.

What is it? How TE is used?

Applications

Textual entailment recognition is used in several NLP tasks:

- *Question Answering*
- Information Extraction
- Information Retrieval
- Text Summarization

and many more.

What is it? How TE is used?

Example:

T: *The technological triumph known as GPS was incubated in the mind of Ivan Getting.*

$$\Downarrow \textit{entails} \tag{1}$$

H: *X invented the GPS*

Textual Entailment in the Community

Recognizing Textual Entailment challenge.
Main Task: given a corpus of T (real data) and a set of H, determine such pairs T-H in which one fragment entails the other.

Textual Entailment in the Community

Recognizing Textual Entailment challenge.
Main Task: given a corpus of T (real data) and a set of H, determine such pairs T-H in which one fragment entails the other.

- compares the performance of TE systems
- launched in 2004 by FBK
- supported by Microsoft Research

Mehdad, Negri, de Souza, Petrova. FBK Participation in the RTE-7 Main Task. Text Analysis Conference, 2011

FBK System for RTE-7

Multifeature system with lexical similarity being the key feature.
An algorithm to compute n-gram match scores for every level of *n*:

- start from 5-grams
- eliminate a string when matched
- repeat for (n-1) level

FBK System for RTE-7

Multifeature system with lexical similarity being the key feature.
An algorithm to compute n-gram match scores for every level of *n*:

- start from 5-grams
- eliminate a string when matched
- repeat for (n-1) level

Extra NLP features: Semantic Roles, Named Entities, Wordnet, Syntactic Dependencies

Conclusion

- TE is an example of how logical notion can be projected to natural language.
- Area of active research.
- Straightforward surface techniques outperform semantic representation approaches...
- ...but clever way of computing lexical similarity should be found to achieve high performance.

Bibliography

- Mehdad, Negri, de Souza, Petrova. FBK Participation in the RTE-7 Main Task. Text Analysis Conference, 2011
- Jia, Huang, Ma, Wan, Xiao. RKUTM Participation at TAC 2010 RTE and Summarization Track. Text Analysis Conference, 2010
- Majumdar, Bhattacharyya. Lexical Based Text Entailment System for Main Task of RTE6. Text Analysis Conference, 2010