

Building Evaluation Dataset for Textual Entailment in Czech

RASLAN 2012, Karlova Studánka

Zuzana Nevěřilová

7. prosince 2012

Intro

Recognizing Textual Entailment

Evaluation Data for RTE

Intro...

Norway's most famous painting, "The Scream" by Edvard Munch, was recovered Saturday, almost three months after it was stolen from an Oslo museum.

Intro...

Norway's most famous painting, "The Scream" by Edvard Munch, was recovered Saturday, almost three months after it was stolen from an Oslo museum.

- ▶ Edvard Munch is a painter
- ▶ Edvard Munch is the author of "The Scream"
- ▶ "The Scream" is a famous painting
- ▶ "The Scream" was stolen
- ▶ "The Scream" was recovered
- ▶ "The Scream" was missing for almost three months
- ▶ "The Scream" was stolen from an Oslo museum
- ▶ Edvard Munch is Norwegian

Intro...

Implicit knowledge exist in text.

[Graesser, 1981] estimates that the ratio explicit:implicit information is about 1:8 which means that the majority of information in texts is not mentioned.

Task: implicit \Rightarrow explicit

Recognizing Textual Entailment: Approaches

A text t entails a hypothesis h if humans reading t will infer that h is most likely true.

[Dagan et al., 2007]

- ▶ bag of words, bag of n-grams
- ▶ graph matching problem for normalized syntactic dependency graphs
- ▶ quasi-logical forms using semantic roles
- ▶ sentences presented as formulae in FOPL



Recognizing Textual Entailment: Approaches

A text t entails a hypothesis h if humans reading t will infer that h is most likely true.

[Dagan et al., 2007]

- ▶ bag of words, bag of n-grams
- ▶ graph matching problem for normalized syntactic dependency graphs
- ▶ quasi-logical forms using semantic roles
- ▶ sentences presented as formulae in FOPL



Recognizing Textual Entailment: Approaches

A text t entails a hypothesis h if humans reading t will infer that h is most likely true.

[Dagan et al., 2007]

- ▶ bag of words, bag of n-grams
- ▶ graph matching problem for normalized syntactic dependency graphs
- ▶ quasi-logical forms using semantic roles
- ▶ sentences presented as formulae in FOPL



Recognizing Textual Entailment: Approaches

A text t entails a hypothesis h if humans reading t will infer that h is most likely true.

[Dagan et al., 2007]

- ▶ bag of words, bag of n-grams
- ▶ graph matching problem for normalized syntactic dependency graphs
- ▶ quasi-logical forms using semantic roles
- ▶ sentences presented as formulae in FOPL



Recognizing Textual Entailment: Approaches

A text t entails a hypothesis h if humans reading t will infer that h is most likely true.

[Dagan et al., 2007]

- ▶ bag of words, bag of n-grams
- ▶ graph matching problem for normalized syntactic dependency graphs
- ▶ quasi-logical forms using semantic roles
- ▶ sentences presented as formulae in FOPL



Building an Evaluation Set

- ▶ reading comprehension tests
 - ▶ Pracovní listy k nácviku porozumění čtenému textu a nápravě čtení – od 2. ročníku ZŠ
 - ▶ PISA (Programme for International Student Assessment)
 - ▶ státní maturita
 - ▶ obecně studijní předpoklady
- ▶ paraphrases generated from corpus texts

Building an Evaluation Set

- ▶ reading comprehension tests
 - ▶ Pracovní listy k nácviku porozumění čtenému textu a nápravě čtení – od 2. ročníku ZŠ
 - ▶ PISA (Programme for International Student Assessment)
 - ▶ státní maturita
 - ▶ obecně studijní předpoklady
- ▶ paraphrases generated from corpus texts

RTE Evaluation

Pracovní listy k nácviku porozumění čtenému textu a nápravě čtení
– od 2. ročníku ZŠ:

T: Zajíc se ani nenadál a koblížek se kutálel dál. Potkal medvěda.
Medvěd křičel: „Já tě sním!“ „Nejez mě, já ti zazpívám: Já
koblížek, koblížek, na másle smažený, na okně chlazený, babičce
jsem utekl, tobě také uteču.“

...

H: Koblížek potkal vlka.

RTE Evaluation

5. třída

T: Mýtus ohledně mrkve a zraku se datuje od časů druhé světové války, kdy se Britské královské letectvo pokoušelo utajit skutečnost, že vyvinulo sofistikovaný radarový systém, s jehož pomocí sestřelovalo německé bombardéry. Vychloubali se, že vysoká přesnost britských pilotů stíhaček je důsledkem konzumace obrovského množství mrkví. Je pravda, že mrkev obsahuje beta-karoten, který je nutný pro dobrý zrak. Tělo přeměňuje beta-karoten na vitamín A a extrémní nedostatek vitamínu A může způsobit slepotu. K dobrému zraku stačí nicméně jen malé množství beta-karotenu a pokud nemáte málo vitamínu A, zrak se vám nezlepší bez ohledu na to, kolik mrkví spořádáte.

H: Mrkev obsahuje beta-karoten.

RTE Evaluation

Státní maturita:

T: Někteří lidé ohrnují nos, když vidí krysu, Bart Weetjens se však usmívá. Tento belgický designer vymyslel způsob, jak mohou často spílaní hlodavci pomáhat řešit globální problém: vyhledávat miny, jichž je v 69 zemích roztroušeno asi 60 milionů. Často se k jejich odhalování používají psi, „já jsem však věděl, že výcvik krys je jednodušší“, říká Weetjens, který tato zvířata v dětství choval. Krysy i málo váží, takže miny pod nimi nevybuchují. V tropických oblastech, kde je uloženo mnoho výbušnin, zůstávají zdravé a je levné je rozmnožovat a chovat.

H: Vyhledávání min je globálním problémem, protože miny jsou roztroušeny ve všech zemích.

RTE Evaluation

PISA:

T: Jen si představte, jak by bylo úžasné dojíždět do práce „virtuálně“ po elektronické dálnici a všechnu práci vyřizovat na počítači nebo po telefonu! Už byste se nemuseli mačkat v přečpaných autobusech a vlacích, ani marnit dlouhé hodiny dojížděním do práce a zase zpátky. Mohli byste pracovat, kde by se vám zachtělo – jen si zkuste představit všechny ty možnosti, které by se tím otevřely!

H: Lidé by měli mít možnost pracovat tolik hodin, kolik chtějí.

Paraphrasing

sentence S_{input} analysis \Rightarrow sentence S_{output} generation

The meaning of S_{input} is the same or *almost* the same as that of S_{output} .

Wonderworks Ltd. constructed the new bridge.

The new bridge was constructed by Wonderworks Ltd.

Wonderworks Ltd. is the constructor of the new bridge.

Paraphrasing Techniques

- ▶ templates:
X painted Y.
Y is the work of X.
- ▶ synonyms:
Children like sweets.
Kids like sweets.
- ▶ hypernyms:
A bomb destroyed the post office on the corner.
A bomb destroyed a building.

Paraphrasing Techniques

- ▶ templates:
X painted Y.
Y is the work of X.
- ▶ synonyms:
Children like sweets.
Kids like sweets.
- ▶ hypernyms:
A bomb destroyed the post office on the corner.
A bomb destroyed a building.

Paraphrasing Techniques

- ▶ templates:
X painted Y.
Y is the work of X.
- ▶ synonyms:
Children like sweets.
Kids like sweets.
- ▶ hypernyms:
A bomb destroyed the post office on the corner.
A bomb destroyed a building.

Paraphrasing Problems

- ▶ monotonicity:

Arafat is targeted for assassination.

Sharon denies Arafat is targeted for assassination.

Arafat byl cílem atentátu.

Šaron popřel, že Arafat byl cílem atentátu.

- ▶ quantifiers:

Dogs barked loudly. \Rightarrow Dogs barked.

No dogs barked loudly. \Rightarrow No dogs barked.

Psi štěkali nahlas. \Rightarrow Psi štěkali.

Žádný pes neštěkal nahlas. \Rightarrow Žádný pes neštěkal.

Paraphrasing Problems

- ▶ monotonicity:

Arafat is targeted for assassination.

Sharon denies Arafat is targeted for assassination.

Arafat byl cílem atentátu.

Šaron popřel, že Arafat byl cílem atentátu.

- ▶ quantifiers:

Dogs barked loudly. \Rightarrow Dogs barked.

No dogs barked loudly. \Rightarrow No dogs barked.

Psi štěkali nahlas. \Rightarrow Psi štěkali.

Žádný pes neštěkal nahlas. \Rightarrow Žádný pes neštěkal.

Building an Evaluation Set, Examples

H-T pairs extracted from corpus:

```
[tag="k1.*c1" ] [] 0,4  
[tag="k2.*nP.*c4.*" ] {0,2} [tag="k1.*nP.*c4" ]  
[tag!="k1.*" ] {0,2} [word=":" ] [tag="k1.*c4" ]  
([word="," ] [tag="k1.*c4" ]) + [word="a|nebo|ani" ]  
[tag="k1.*c4" ] within <s/>
```

T: Kniha má tři části: úvod, stať a závěr.

H: Úvod je část knihy.

H: Stať je část knihy.

H: Závěr je část knihy.

Building an Evaluation Set, Examples

H-T pairs extracted from corpus:

```
[tag="k1.*c1.*"]{1,2} ([word=",")? [lemma="jinak"]  
([lemma="těž|taky|také"])? [lemma="zvaný"]  
([lemma="těž|taky|také"])? [tag="k1.*c1.*"] {1,3}  
within <s/>
```

T: ...Noc živých mrtvých (režie George A. Romero, 1969), jemuž vévodí oživlé mrtvoly, jinak zvané zombie.

H: Oživlým mrtvolám se taky říká zombie.

Building an Evaluation Set, Examples

H-T pairs extracted from corpus:

```
[tag="k2.*c1.*"]{0,3} [tag="k1.*c1.*"] [lemma="být"]  
[tag="k5e.a.mNg.n."& lemma!="vybavit"&  
lemma!="opatřit"& lemma!="zvolit"& lemma!="vyhlásit"&  
lemma!="dát"] [tag="k1.*c7.*"] [tag=".*c2.*"]{0,8}  
within <s/>
```

T: Zatímco Grossovi během projevu nikdo ani jednou nezatleskal, Škromachův projev byl přerušen aplausem hned dvakrát.

H: aplaus přerušil Škromachův projev

Building an Evaluation Set, Examples

H-T pairs extracted from corpus:

```
[tag="k2.*c1.*"]{0,3} [tag="k1.*c1.*"] [lemma="být"]  
[tag="k5e.a.mNg.n."& lemma!="vybavit"&  
lemma!="opatřit"& lemma!="zvolit"& lemma!="vyhlásit"&  
lemma!="dát"] [tag="k1.*c7.*"] [tag=".*c2.*"]{0,8}  
within <s/>
```

T: Dívka byla uškrcena řemínkem od kabelky.

H: řemínek uškrtil dívku

Building an Evaluation Set

Annotation using Sysel

Annotation instructions:

annotate YES iff:

- ▶ H makes sense (syntactic errors in number or gender rare but possible)
- ▶ human would say that H is entailed by T

annotate NO iff:

- ▶ H does not make sense
- ▶ human would say that H is not entailed by T
- ▶ sentence constituents in H differs from that in T

Building an Evaluation Set

Annotation using Sysel

Annotation instructions:

annotate YES iff:

- ▶ H makes sense (syntactic errors in number or gender rare but possible)
- ▶ human would say that H is entailed by T

annotate NO iff:

- ▶ H does not make sense
- ▶ human would say that H is not entailed by T
- ▶ sentence constituents in H differs from that in T

Building an Evaluation Set

Annotation using Sysel

Annotation instructions:

annotate YES iff:

- ▶ H makes sense (syntactic errors in number or gender rare but possible)
- ▶ human would say that H is entailed by T

annotate NO iff:

- ▶ H does not make sense
- ▶ human would say that H is not entailed by T
- ▶ sentence constituents in H differs from that in T

Building an Evaluation Set

Annotation using Sysel

Annotation instructions:

annotate YES iff:

- ▶ H makes sense (syntactic errors in number or gender rare but possible)
- ▶ human would say that H is entailed by T

annotate NO iff:

- ▶ H does not make sense
- ▶ human would say that H is not entailed by T
- ▶ sentence constituents in H differs from that in T

Building an Evaluation Set

Annotation using Sysel

Annotation instructions:

annotate YES iff:

- ▶ H makes sense (syntactic errors in number or gender rare but possible)
- ▶ human would say that H is entailed by T

annotate NO iff:

- ▶ H does not make sense
- ▶ human would say that H is not entailed by T
- ▶ sentence constituents in H differs from that in T

Building an Evaluation Set: Annotation Pitfalls

Difference in sentence constituents in H and T:

T: ... že tyto věci podivných tvarů shromažďují nadpřirozené bytosti: skřítkové, víly, draci a ďáblové

H: skřítek je nadpřirozená bytost

H: skřítek je bytost

T: Pro zpestření života hotelových hostů se na pláži denně konaly sportovní akce: aerobik, tancování, volejbal, kuželky a fotbal.

H: aerobik je sportovní akce

H: aerobik je akce

Building an Evaluation Set: Annotation Pitfalls

Difference in sentence constituents in H and T:

T: ... že tyto věci podivných tvarů shromažďují nadpřirozené bytosti: skřítkové, víly, draci a ďáblové

H: skřítek je nadpřirozená bytost

H: skřítek je bytost

T: Pro zpestření života hotelových hostů se na pláži denně konaly sportovní akce: aerobik, tancování, volejbal, kuželky a fotbal.

H: aerobik je sportovní akce

~~H: aerobik je akce~~

sportovní akce není akce

Building an Evaluation Set: Annotation Pitfalls

Exclusivity

T: Nejznámější Češi se nazývají: Kožený, Krejčíř a Pitr

H: Kožený je nejznámější Čech

H: Krejčíř je nejznámější Čech

H: Pitr je nejznámější Čech

Building an Evaluation Set: Numbers

CZES: 465,102,710 tokens

Enumeration: 738 hypotheses

Passive: 12.573 hypotheses

Aliases: 26 hypotheses



Dagan, I., Roth, D., and Zanzotto, F. M. (2007).

Tutorial notes.

In 45th Annual Meeting of the Association of Computational Linguistics. The Association of Computational Linguistics.



Graesser, A. (1981).

Prose comprehension beyond the word.

Springer-Verlag.