

RTE-5 Planning Session

Ido Dagan

Bernardo Magnini

Luisa Bentivogli

Hoa Trang Dang

Danilo Giampiccolo

Main Issues

- Future of the *MAIN/SEARCH* tasks
 - Potential for IE/KBP-based RTE dataset
 - Venue for Inference Resources
 - Knowledge & modules
 - Towards Specialized Datasets: preliminary investigation
- ➔ open discussion and audience feedback

Roadmap for RTE Datasets

- Within several years – create datasets that reflect realistic application usage
 - Correspond to useful application scenarios
 - Reflect true distribution of entailments (in each application)
- Gradually cover prominent applications, run each scenario for a few years
- Direct RTE system developers to meet application-performance needs
- Move away from the traditional main task?

Search Task for SUM

How can we make the RTE Search scenario as close as possible to a realistic SUM scenario?

1. H's based on real texts snippets rather than on artificial SCUs
2. Update scenario: H's based on text snippets from SUM collection B and corpus=A
3. H's from an automatic summary and T's are the rest of the summary (detecting redundancy)

We need people from summarization to help in the task design!

IE/KBP-based Dataset?

- Hs – template statements, with typed variables, corresponding to the slot filling task
 - *X<organization> employs Y<person>*
 - *IBM employs <person>*
- Texts taken from KBP corpus or Wikipedia pages
- Can cover flexibly many relation/slot types
- Test relations are not known in advance
 - Aim at on-demand/unsupervised slot filling scenarios, either without or with minimal training, leveraging the broad coverage potential of RTE technology

Venue for inference resources

- Motivation: sharing and reuse of resources
- Call for resource contributions (which date?)
- Formal lists of resources and announcement
- Ablation Tests
- Results published
 - Credits for resource contributors

Specialized Datasets

- Different linguistic phenomena are involved in textual entailment (e.g. lexical, syntactic, discourse) and interact in a complex way
- it is difficult to measure a substantial impact of specific modules on the RTE datasets:
 - sparseness (low frequency) of the single phenomena
 - impossibility to isolate each phenomenon (i.e. other system components can influence the final judgment)

Specialized Datasets: 1st attempt

Creation of specialized TE datasets made of *monothematic* T-H pairs, i.e. pairs in which a certain phenomenon underlying the entailment relation is highlighted and isolated

Main Issues:

- is it possible to clearly identify and isolate the linguistic phenomena underlying the entailment relation?
- how to categorize phenomena?
- how easy/difficult it is to create balanced datasets with respect to the distribution of positive and negative examples?

Positive Examples

TEXT(S)		HYPOTHESES		LINGUISTIC PHENOMENA
T_{orig}	British writer <u>Doris Lessing</u> , recipient of the <u>2007 Nobel Prize in Literature</u> , has said in an interview that the terrorist attack on September 11 "wasn't that terrible" when compared to attacks the Irish Republican Army (IRA) made on Britain. [...]	H_{orig}	Doris Lessing won the Nobel Prize in Literature in 2007.	-apposition -verbalization -synonymy -modifier
		H1	Doris Lessing <u>is</u> the recipient of the 2007 Nobel Prize in Literature.	apposition
		H2	Doris Lessing <u>received</u> the 2007 Nobel Prize in Literature.	verbalization
		H3	Doris Lessing, <u>winner</u> of the 2007 Nobel Prize in Literature, has [...]	synonymy

Negative Examples: Contradictions

TEXT		HYPOTHESES		LINGUISTIC PHENOMENA
T _{orig}	<u>Mexico's new president, Felipe Calderon</u> , seems to be doing all the right things in cracking down on Mexico's drug [...]	H _{orig}	Felipe Calderon is the outgoing President of Mexico.	-antonym - apposition - genitive
		H ₁	Mexico's outgoing president, Felipe Calderon, seems to [...]	antonym

NB: *apposition* and *genitive* phenomena DO preserve entailment

Problems:

- contradictions are only the 15% of the DataSet
- the variety of phenomena in negative examples is reduced with respect to the positive pairs

Negative Examples: Unknown

- H contains additional information not inferable by T

T: *FIFA has begun disciplinary proceedings against Marco Materazzi after Zinedine Zidane was sent off for headbutting Materazzi in the chest. [...]*

H: *Materazzi insulted Zidane.*

- It is not possible to identify (and isolate) linguistic phenomena responsible for “non-entailment”

Negative Examples: Current Option

- generating the negative examples by modifying parts of the positive examples, to obtain a contradiction due to the same phenomenon

TEXT(S)		HYPOTHESES		LINGUISTIC PHENOMENA
T	British writer <i>Doris Lessing</i> , recipient of the <i>2007 Nobel Prize in Literature</i> , has said in an interview that [...]	H1 Positive	Doris Lessing <u>received</u> the 2007 Nobel Prize in Literature.	verbalization
		H2 Negative	Doris Lessing <u>refused</u> the 2007 Nobel Prize in Literature.	verbalization ?antonym

Suggestions? Comments?

Thank you

See you all next year!