Determining is-a Relationships for Textual Entailment

Vlad Niculae · vlad@vene.ro · Université de Franche Comté, University of Bucharest
Octavian Popescu · popescu@fbk.eu · Fondazione Bruno Kessler, Trento

Textual Entailment and the RTE Corpora

Given a pair of text (T) and hypothesis (H),
would a human reading T believe that H is true?
H is always well-formed and logical in the RTE corpora.
What if H is not carefully produced by the user?
If H candidates are massively machine generated,
robustness and **precision** are needed.

Common Approaches:
Handcrafted rules (Iftene & Moruz, 2009): ~70% precision on RTE5
Machine Learning: competitive accuracy, usually don't report precision

Recognizing is-a Textual Entailment

Subset of RTE{3,4,5} where
H ~ [[X is α]]
Extract attributes with syntax and pattern matching
([[X is α]], [[α/PMOD X]], [[X is known as Y]])

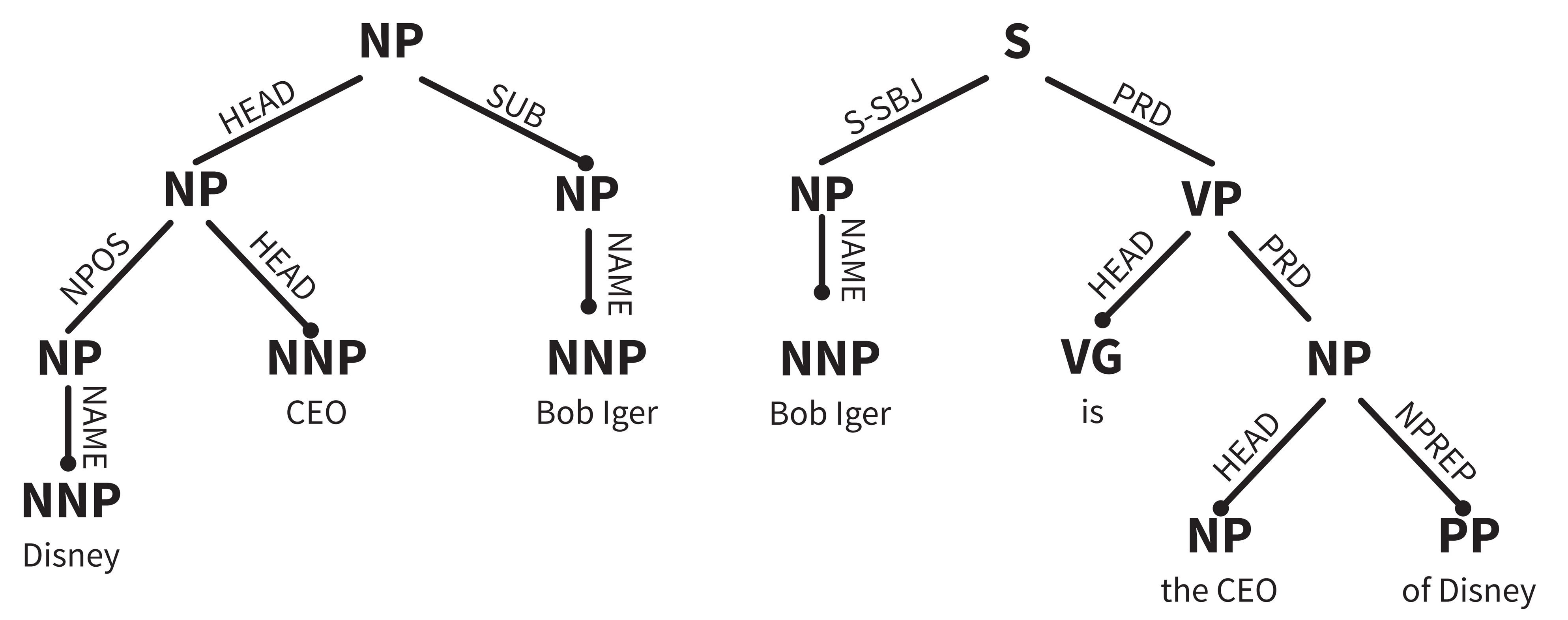
1. Match: H ~ [[X **be** α]]
2. Find entities X_1, X_2, ... in T and their attributes
3. Identify the X_i that corefer, merge their attributes
4. Match X against the result, check for α

Example:
T: For weeks now, I've been hearing chatter that Disney was close to doing a distribution deal with Hulu. [...] And accompanying said chatter was this refrain: Why? The puzzlement comes from video players who don't work at NBC, Fox or Hulu, and who can't see the upside in Disney CEO Bob Iger throwing in his lot with Hulu.

H: Bob Iger is the CEO of Disney.

GLARF Pattern Matching

Grammatical and Logical Argument Representation Format
Applies normalizations based on syntax (Zellig & Harris, LFG, HPSG) and NLP resources (COMLEX, NOMLEX, PTB, PropBank) to reduce variation.



Results

Baseline 1: lexical overlap, linear SVM
Baseline 2: lexical, syntactic and NE, overlap and set difference, linear SVM
Baseline 3: Tree kernel SVM on subtrees

System attributes:
S/N : match synonyms or not
B/A: match entities at word boundaries or anywhere

# +/-	RTE-3′ 104 / 91				RTE-4′ 90 / 102				RTE-5′ 134 / 131			
	A	P	R	F	A	P	R	F	A	P	R	F
BL1	.69	.71	.69	.70	.51	.49	.35	.40	.57	.56	.84	.67
BL2	.74	.71	.87	.78	.62	.62	.53	.57	.59	.59	.82	.68
BL3	.56	.61	.45	.52	.57	.53	.53	.53	.59	.59	.65	.61
NB	.57	.96	.21	.35	.62	.90	.21	.34	.44	.96	.18	.30
NA	.56	.92	.21	.34	.62	.88	.23	.37	.45	.96	.20	.34
SB	.56	.88	.22	.35	.62	.82	.26	.39	.44	.82	.21	.34
SA	.57	.86	.24	.38	.64	.81	.29	.43	.45	.78	.26	.39

Hypothesis Scrambling

It seems that in RTE, positive entailment is correlated with the sharing of entities between T and H.
This makes overlap-based methods unrealistically good.

For each positive pair $(T(X_1, X_2, \ldots), H(X \text{ is } \alpha))$,
generate negative pairs $(T(X_1, X_2, \ldots), H(X_i \text{ is } \alpha))$

E.g. for the text above:
✓ (Bob Iger is the CEO of Disney.)
✗ I is the CEO of Disney.
✗ Chatter is the CEO of Disney.
✗ Disney is the CEO of Disney.
✗ NBC is the CEO of Disney.
✗ Fox is the CEO of Disney.

Accuracy:

#	RTE3 673	RTE4 315	RTE5 418
BL1	.50	.75	.20
BL2	.15	.32	.18
BL3	.65	.60	.54
NB	.90	.95	.95
NA	.90	.94	.94
SB	.84	.91	.87
SA	.79	.85	.80