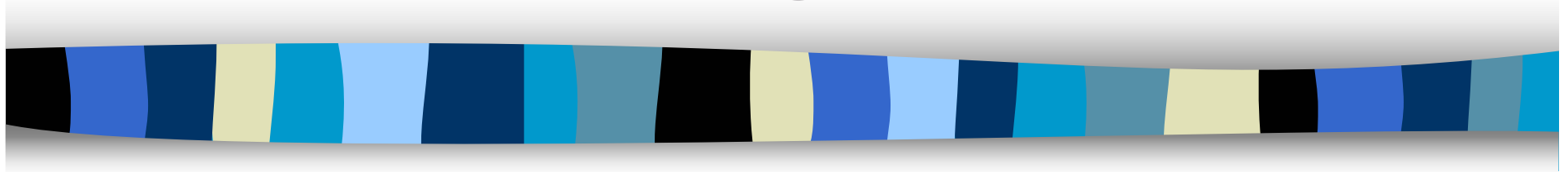


Recognizing Textual Relatedness with Predicate-Argument Structures



Rui Wang
Joint work with Dr. Yi Zhang

Saarland University
Germany



Introduction of Textual Entailment

- What is *Textual Entailment* (Chierchia and McConnell-Ginet, 2001)?
 - Text: *We are having the conference, EMNLP 2009, in Singapore now.*
 - Hypothesis: *We have plenty of durians to eat.*
- What is *Recognizing Textual Entailment* (RTE) (Dagan et al., 2006)?
 - Yearly challenge, datasets, 20+ participants, ...
 - state-of-the-art performance: ~**60%** (2-way)



From 2-Way to 3-Way

- Annotation
 - Yes → Entailment
 - No → Contradiction / Unknown
- Results
 - RTE-4 (3-way): 0.51; RTE-4 (2-way): 0.57
 - RTE-3 (2-way): 0.61



An Example

- Text: *At least five people have been killed in a head-on train collision in north-eastern France, while others are still trapped in the wreckage. All the victims are adults.*
- Hypothesis: *A French train crash killed children.*
-



An Example

- Text: *At least five people have been **killed** in a head-on **train collision** in north-eastern **France**, while others are still trapped in the wreckage. All the victims are adults.*
- Hypothesis: A **French train crash** killed children.
- **Contradictive but relevant!**



Strategies

- Traditional 2-way classification
 - Split E cases first: $ECU \rightarrow E/CU$
- Contradiction recognition (de Marneffe et al., 2008)
 - Split C cases first: $ECU \rightarrow C/EU$
- Others
 - Three-way classification: $ECU \rightarrow E/C/U$
 - Split U cases first: $ECU \rightarrow U/EC$



Baseline

- RTE-4 Dataset
 - 500 E, 150 C, 350 U
 - NaiveBayes classifier, 10-fold CV
 - BoW + SynDep features (Wang and Neumann, 2007)

Three-Way	Two-Stage		
<i>E/C/U</i>	<i>E/CU → E/C/U</i>	<i>C/EU → C/E/U</i>	<i>U/EC → U/E/C</i>



Outline

- Textual Relatedness
 - Related work
 - Definition
- Predicate-Argument Structure (PAS) Extraction
 - Related work
 - Recognizing Textual Relatedness (RTR)
- Experiments
 - Results
- Summary



RTE vs. RTR

- RTE
 - Direct three-way classification (e.g. Agichtein et al., 2009); different rules simultaneously (Clark and Harrison, 2009)
 - Contradiction recognition (de Marneffe et al., 2008)
- Alignment
 - Phrased-based and dependency-graph-based (Pado et al., 2009)
 - Ontology-based (Siblini and Kosseim, 2009)
 - Dependency-path-based (Wang and Neumann, 2007)



Textual Relatedness

- If **H** is fully relevant to part of **T**, **H** is *semantically related* to **T**.
- Relatedness
 - (Weaker than) Similarity (surface string, semantic, etc.)
 - (Stronger than) Co-occurrence (distributionally or ontologically)

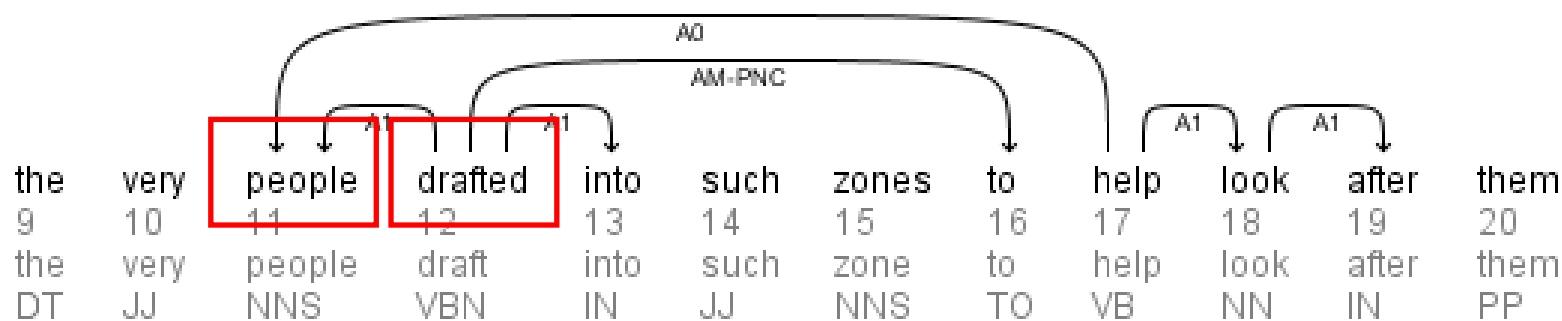


SRL / PAS for RTE

- Semantic role labeling
 - The limited coverage of the verb frames or predicates;
 - The undetermined relationships between two frames or predicates;
 - The unsatisfying performance of an automatic SRL system.
- The CoNLL shared task (2008, 2009)
 - Syntactic and semantic dependency learning
 - State-of-the-art performance: 70~80%

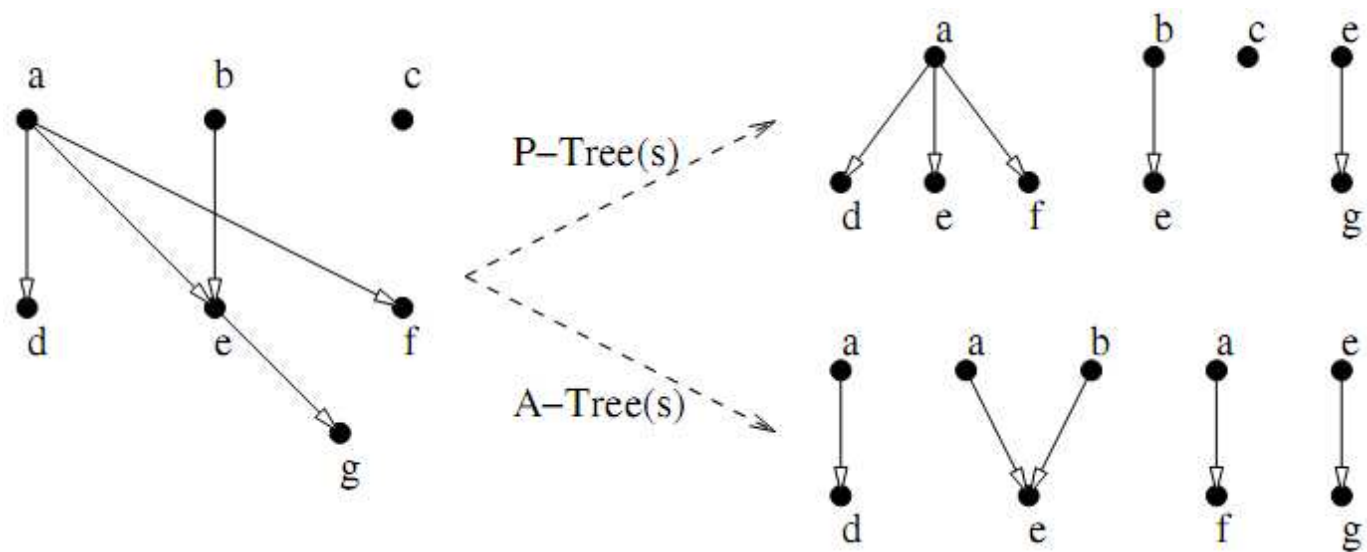
Predicate-Argument Graph (PAG)

- The SRL system (Zhang, Wang&Uszkoreit, 2008)
 - E.g. *Children in post-conflict areas are being abused by the very people drafted into such zones to help look after them ...*

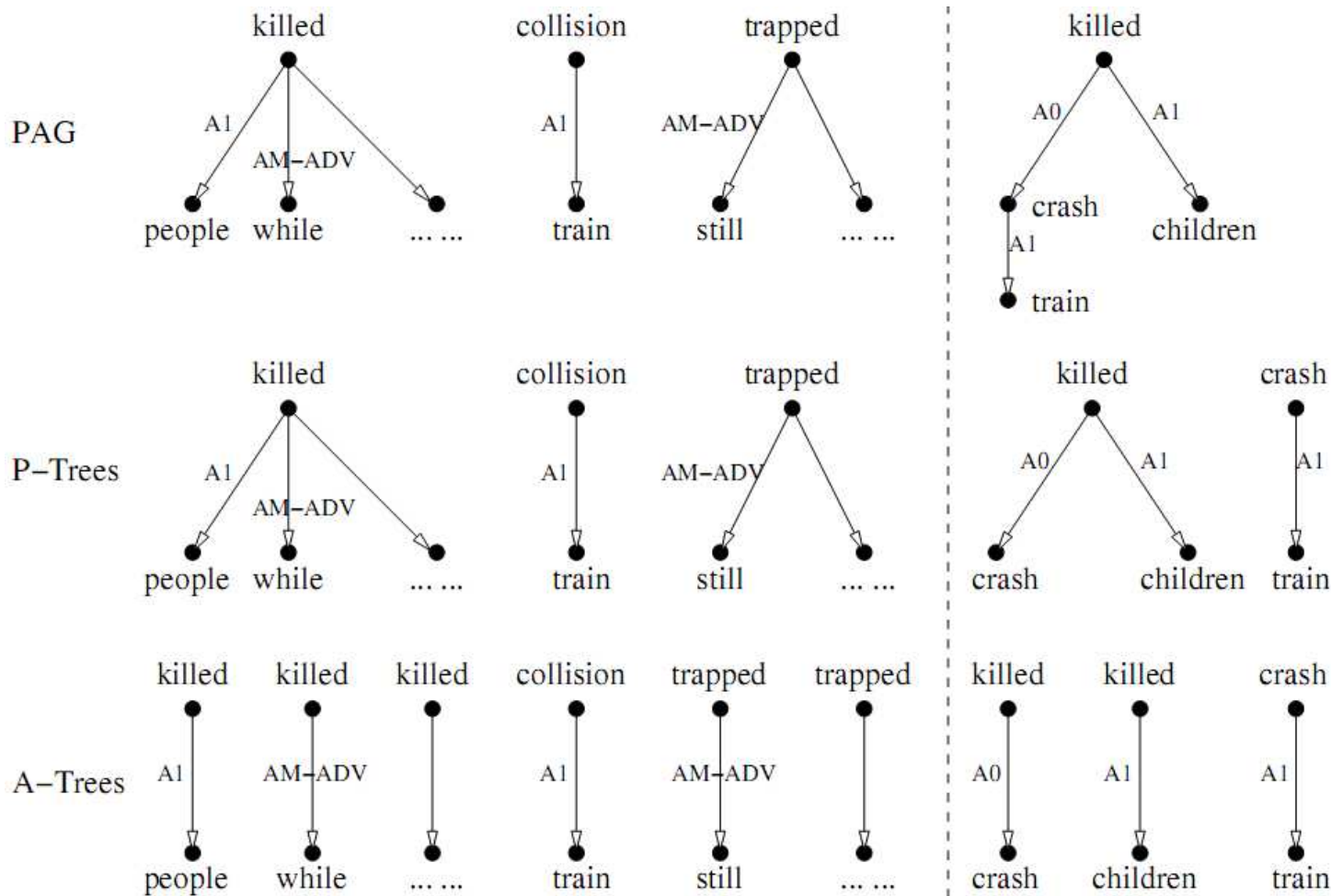


Decomposition of PAG

- Predicate Trees (P-Trees)
- Argument Trees (A-Trees)



An Example





Equations

$$R(T, H) = \max_{1 \leq i \leq r, 1 \leq j \leq s} \{R(\text{Tree}_{T_i}, \text{Tree}_{H_j})\}$$

$$R(\text{Tree}_T, \text{Tree}_H) = \min_{1 \leq i \leq n, 1 \leq j \leq m} \{R(\langle P_T, D_{T_i}, A_{T_i} \rangle, \langle P_H, D_{H_j}, A_{H_j} \rangle)\}$$

$$R(\langle P_T, D_T, A_T \rangle, \langle P_H, D_H, A_H \rangle) = \begin{cases} \text{Full} & R(P_T, P_H) = R(D_T, D_H) = R(A_T, A_H) = 1 \\ \text{NotFull} & R(P_T, P_H) = R(D_T, D_H) = 1 \\ \text{Other} & \text{Otherwise} \end{cases}$$



Lexical Semantic Resources

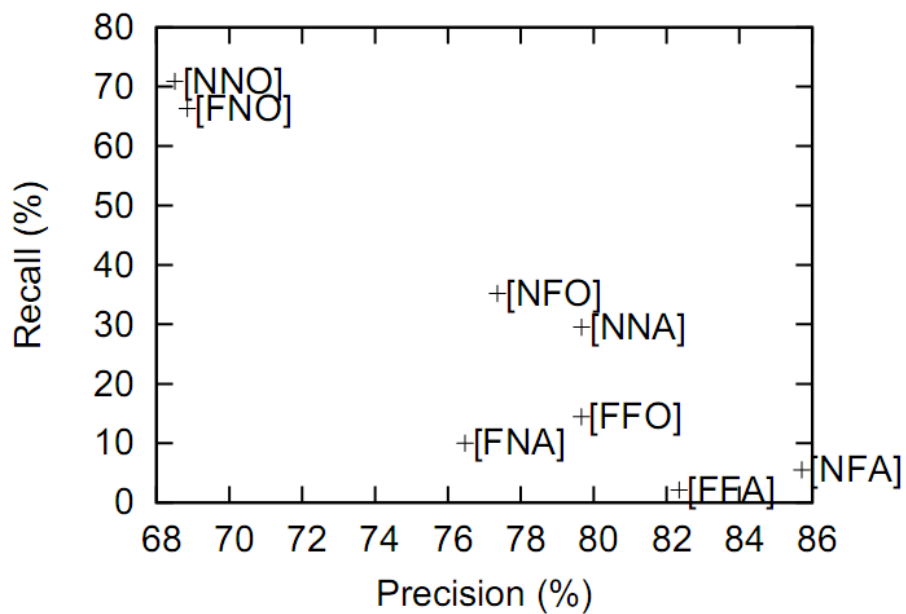
- String matching of lemmas
- Predicate
 - VerbOcean (Chklovski and Pantel, 2004)
 - Normalized Google Distance (NGD) (Cilibrasi and Vitanyi, 2007)
- Argument
 - WordNet: synonym, hypernym, hyponym, antonym
 - NGD



Experiments

- System
 - Baseline: BoW+Syn
 - Main: Baseline + PAS
- Datasets
 - RTE-4 test set
 - 500 entailment T-H pairs (E); 150 contradiction pairs (C); 350 unknown pairs (U)
 - RTE-3 dev+test set
 - 822 E pairs; 161 C pairs; 617 U pairs

Results



	Precision	Recall
[FFA]	0.8235	0.0215
[NFA]	0.8571	0.0554
[FNA]	0.7647	0.1
[FFO]	0.7966	0.1446
[NNA]	0.7967	0.2954
[NFO]	0.7736	0.3523
[FNO]	0.6885	0.6631
[NNO]	0.685	0.7092
[---]	0.65	1

Systems	The First Stage		
	Baseline2	SRL+Baseline2	SRL
RTE-3 [NNA]	59.50%	60.56%(1.78%↑)	70.33%
RTE-4 [NFO]	67.10%	70.20%(4.62%↑)	79.67%

Results (cont.)

- The final results

Systems	Baseline1	Baseline2	SRL+Baseline2
Data Sets	Three-Way	Two-Stage	Two-Stage
RTE-3 [NNA]	52.19%	52.50%	53.69%(2.87%↑)
RTE-4 [NFO]	53.20%	54.20%	56.60%(6.39%↑)

- Lexical resources

Data Sets	SRL+Baseline

SRL+Baseline - VO	SRL+Baseline - NGD	SRL+Baseline - WN



Summary

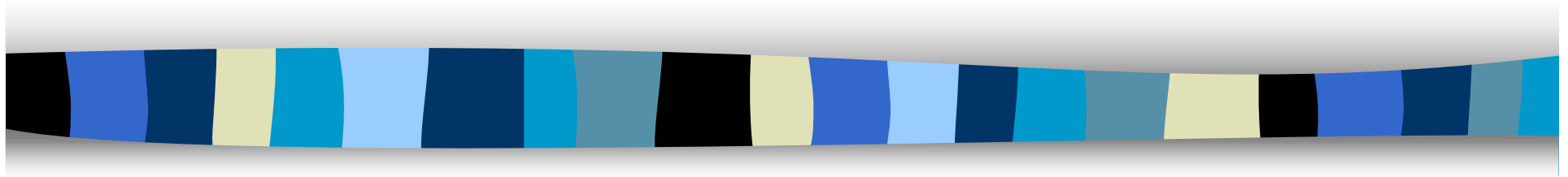
- Strategy
 - 2-stage binary classification for 3-way RTE
- Approach
 - Textual relatedness
 - Use PAS to measure it
- Result
 - Improved (combination)
 - Different lexical resources contribute differently



Future Work

- Combined with other methods/systems
- Two styles of alignment
 - Predicate (Dinu and Wang, 2009)
 - Argument (Callison-Burch, 2008)
- Entailment vs. Contradiction
 - Fine-grained usage of PAS

Thank you!



Questions?