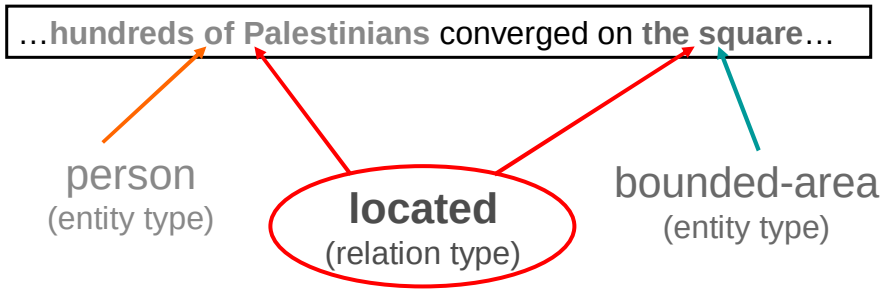
A Systematic Exploration of the Feature Space for Relation Extraction

Jing Jiang & ChengXiang Zhai

Department of Computer Science
University of Illinois, Urbana-Champaign

What Is Relation Extraction?

What Is Relation Extraction?



Existing Methods

- Rule-based [Califf & Mooney 98]
- Generative-model-based [Miller et al. 00]
- Discriminative-model-based
  - Feature-based [Zhou et al. 05]
  - Kernel-based [Bunescu & Mooney 05b] [Zhang et al. 06]

Feature-Based Methods

…**hundreds of Palestinians**$_{arg1}$ converged on **the square**$_{arg2}$…

located

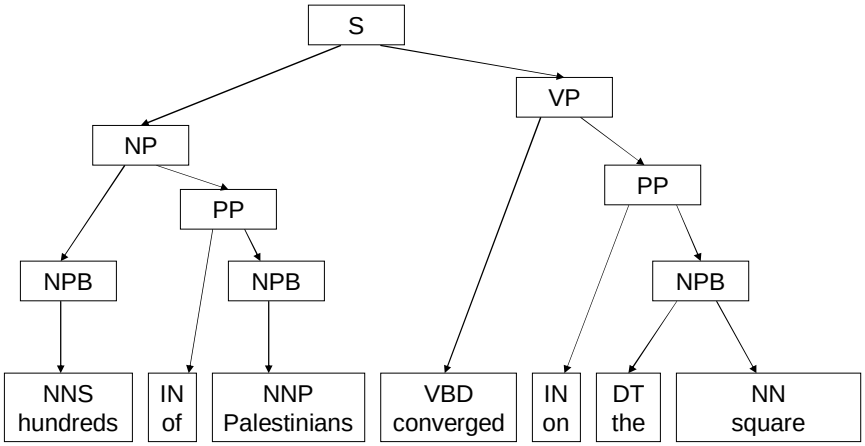
- Entity info
  - arg_1 is a Person entity & arg_2 is a Bounded-Area entity
- POS tagging
  - there is a pr… …rg$_2$
- Syntactic pa…
  - arg_2 is inside a prepositional phrase following arg_1
- Dependency parsing
  - arg_2 is dependent on a preposition, which in turn is dependent on a verb

Other features?

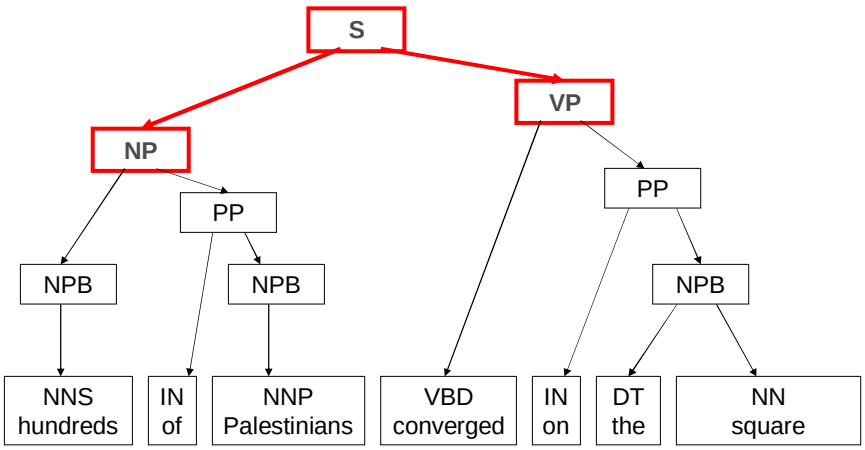
Kernel-Based Methods

- Define a kernel function to measure the similarity between two relation instances
- Convolution kernels
  - Defined on sequence or tree representation of relation instances
  - Corresponding to a feature space, where features are sub-structures such as sub-sequences and sub-trees

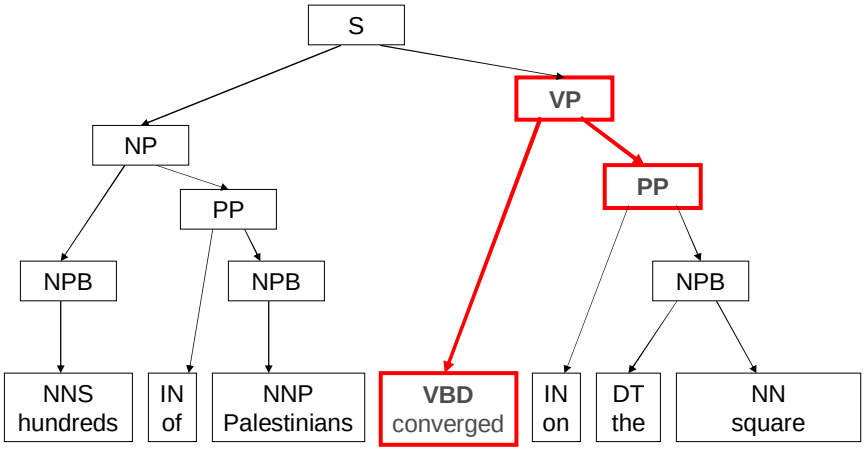
Convolution Tree Kernel (sub-tree features)



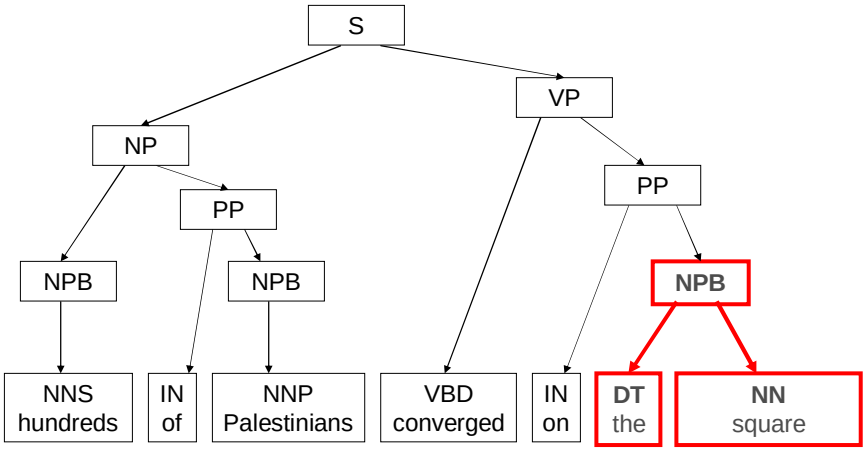
Convolution Tree Kernel (sub-tree features)



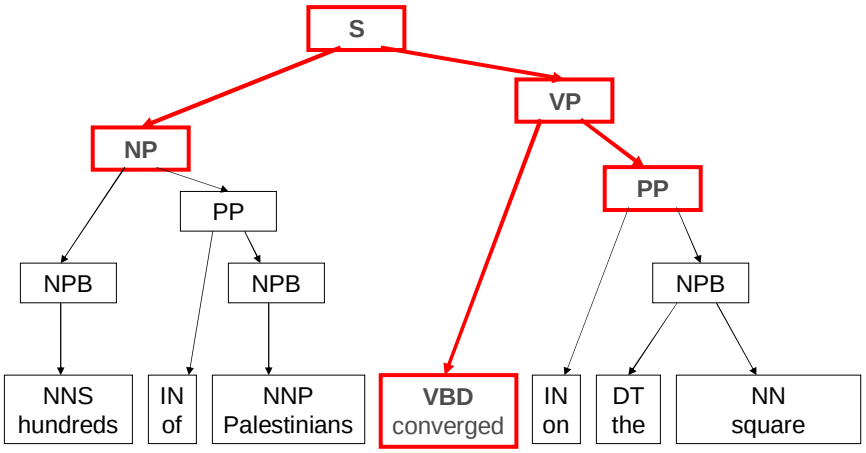
Convolution Tree Kernel (sub-tree features)



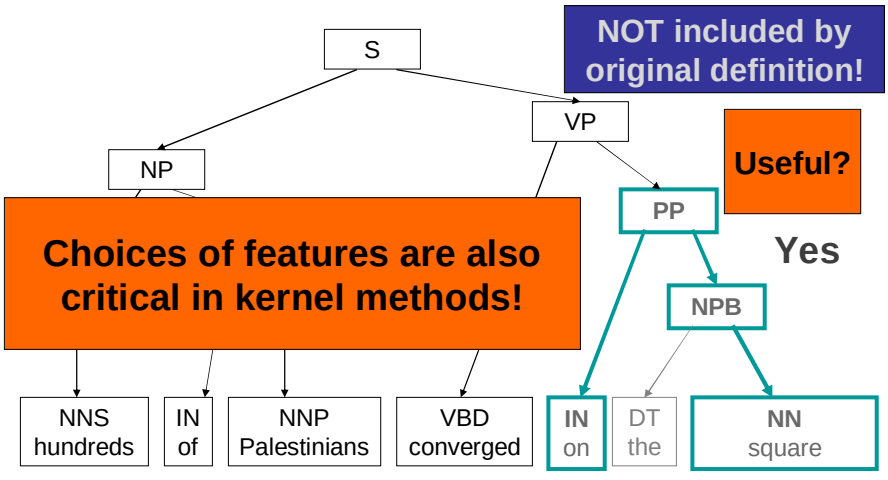
Convolution Tree Kernel (sub-tree features)



Convolution Tree Kernel (sub-tree features)



Convolution Tree Kernel (sub-tree features)



Is it possible to define the complete set of potentially useful features?

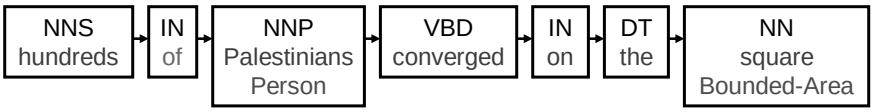
Outline of Our Work

- Defined a graphic representation of relation instances
- Presented a general definition of features
- Proposed a bottom-up search strategy to explore the feature space
- Evaluated different types of features

A Graphic Representation of Relation Instances

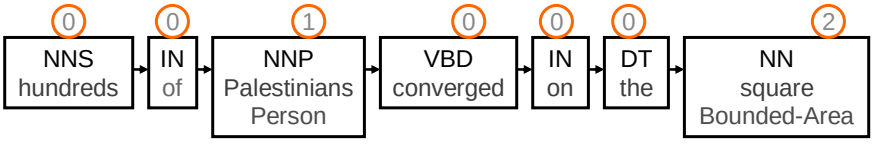
- Each node can have multiple labels
  - Word, POS tag, entity type, etc.

A Graphic Representation of Relation Instances



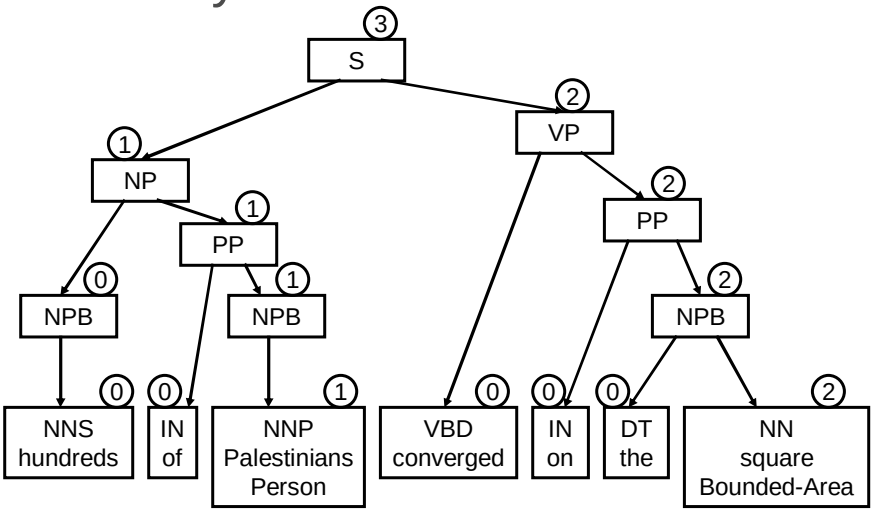
- Each node can have multiple labels
  - Word, POS tag, entity type, etc.
- Each node has an argument tag set to 0, 1, 2, or 3

A Graphic Representation of Relation Instances

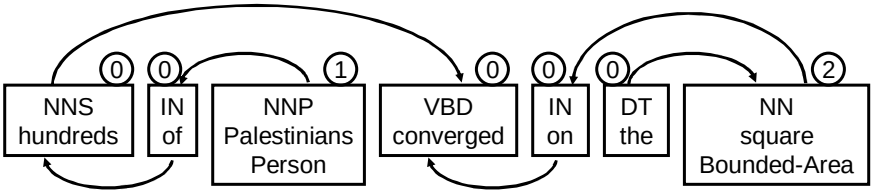


- Each node can have multiple labels
  - Word, POS tag, entity type, etc.
- Each node has an argument tag set to 0, 1, 2, or 3

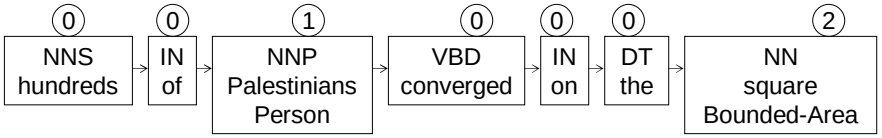
Graphic Representation Based on Syntactic Parse Trees



Graphic Representation Based on Dependency Parse Trees

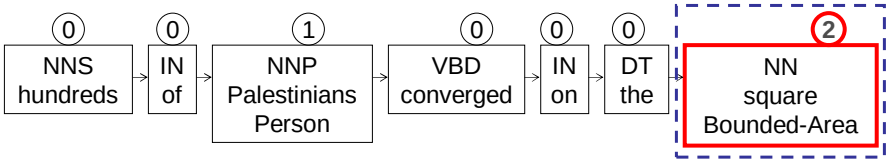


A General Definition of Features



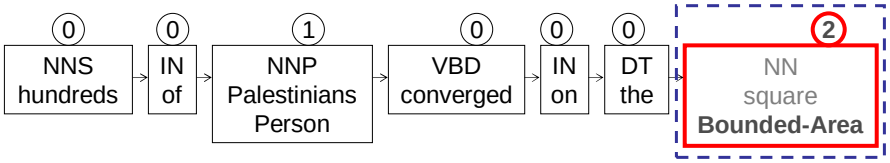
- Sub-graphs

A General Definition of Features



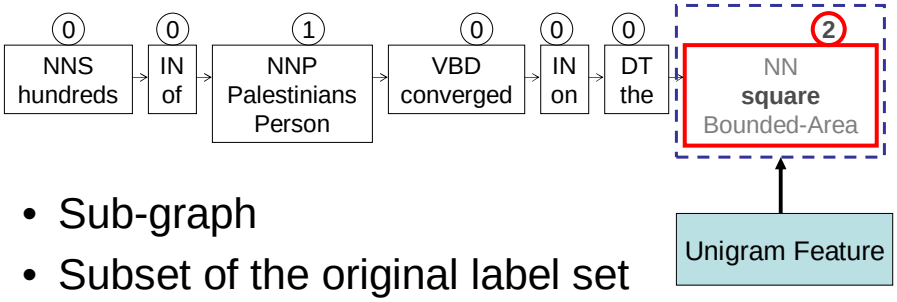
- Sub-graph
- Subset of the original label set

A General Definition of Features



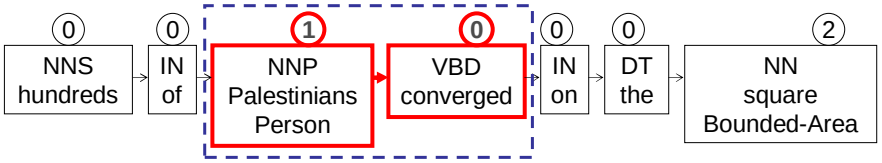
- Sub-graph
- Subset of the original label set

A General Definition of Features

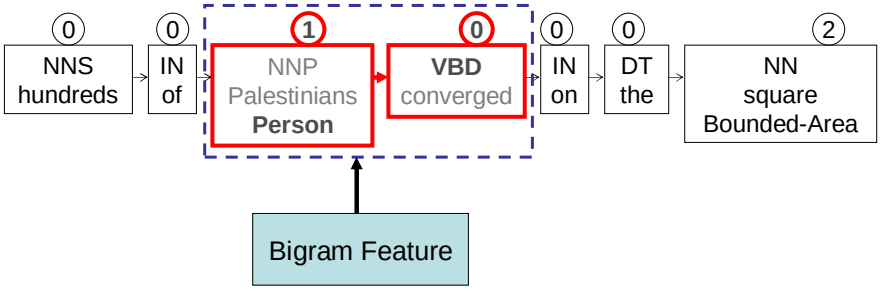


- Sub-graph
- Subset of the original label set

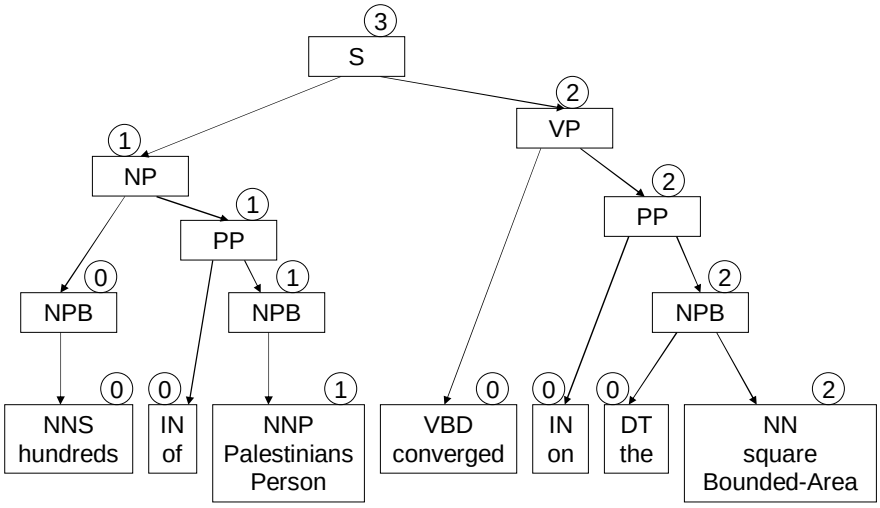
A General Definition of Features



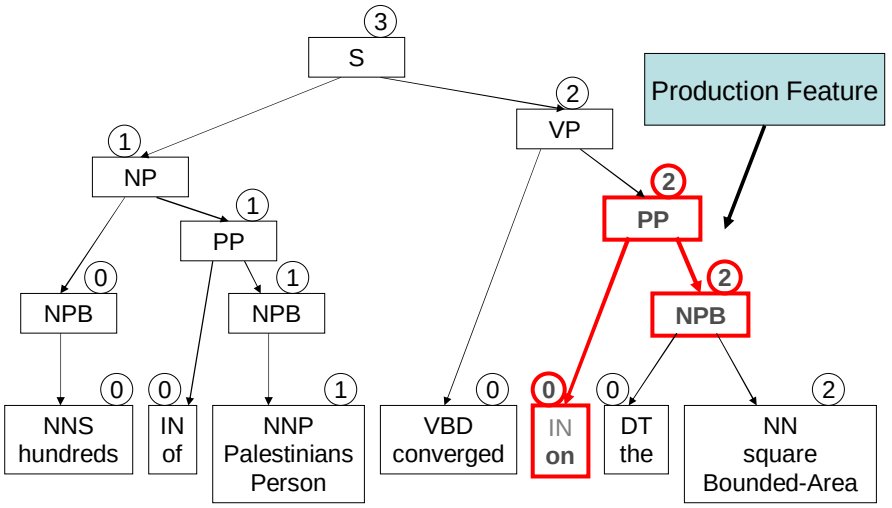
A General Definition of Features



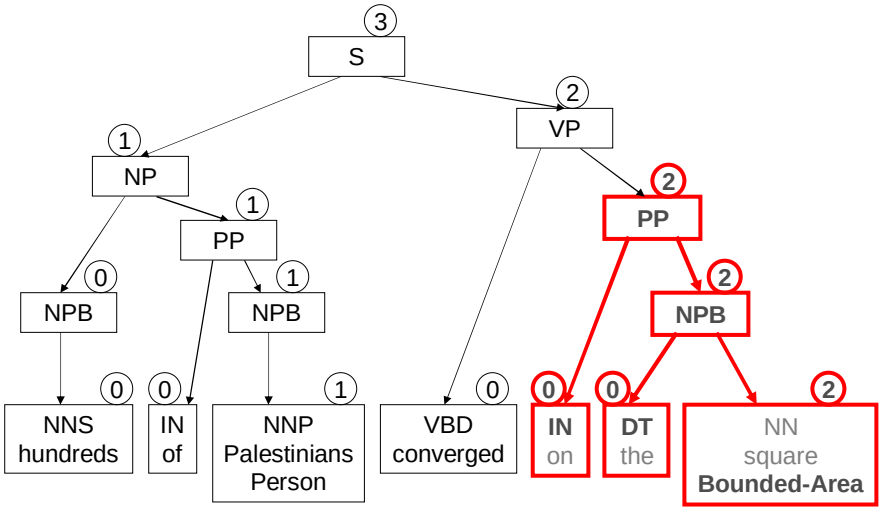
More Examples



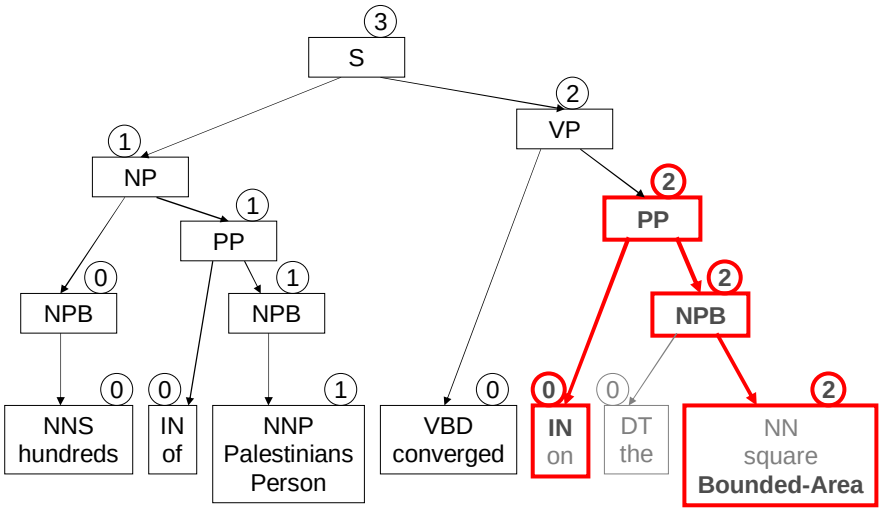
More Examples



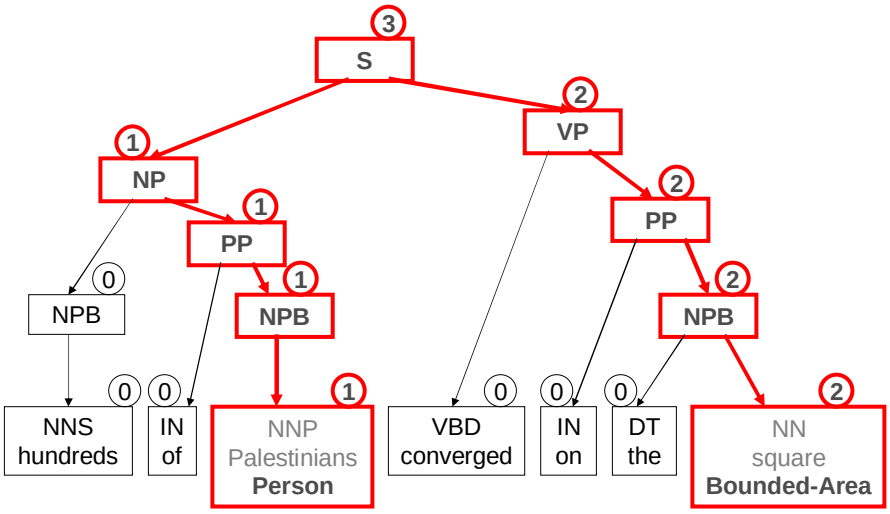
More Examples



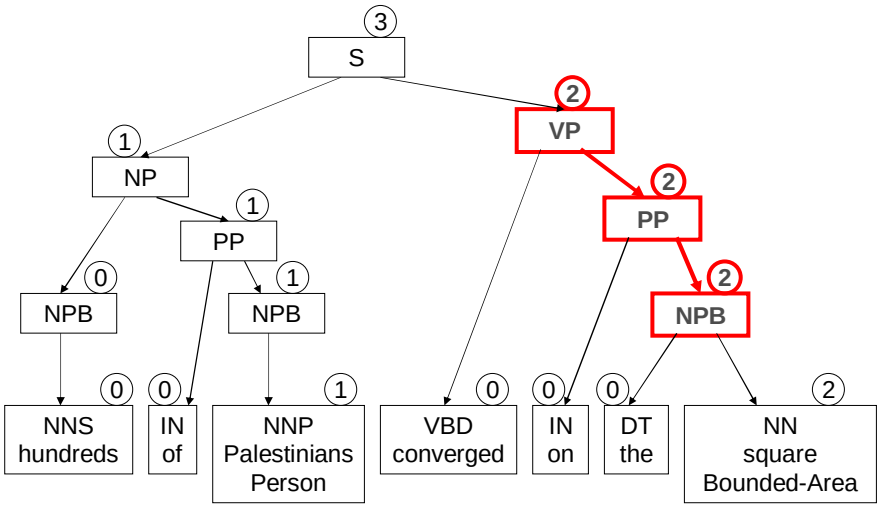
More Examples



More Examples

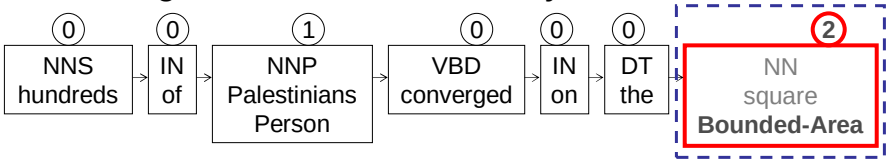


More Examples



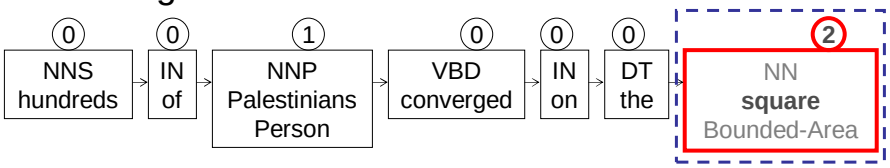
Coverage of the Feature Definition

- Entity attributes [Zhao & Grishman 05] [Zhou et al. 05]
  - Unigram features with entity attributes



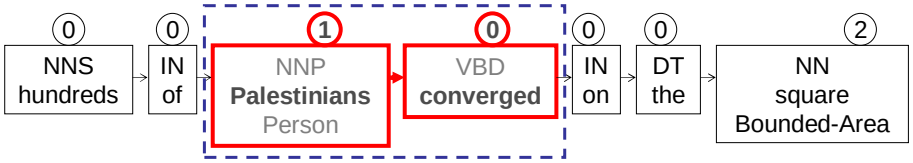
Coverage of the Feature Definition

- Bag-of-word features [Zhao & Grishman 05] [Zhou et al. 05]
  - Unigram features with words



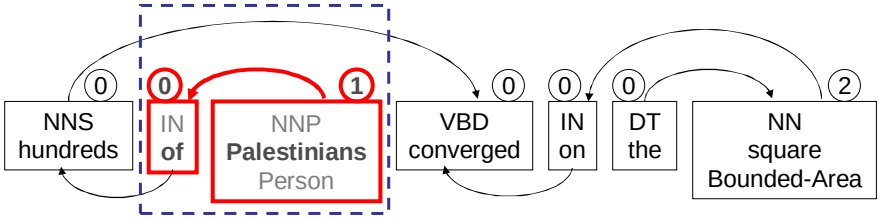
Coverage of the Feature Definition

- Bigram features [Zhao & Grishman 05]
  - Bigram features with words



Coverage of the Feature Definition

- Grammar production features [Zhang et al. 06]
  - Production features
- Dependency relation and dependency path features [Bunescu & Mooney 05a] [Zhao and Grishman 05] [Zhou et al. 05]
  - Bigram and n-gram features with words



Exploring the Feature Space

- We consider three feature subspaces:
  - Sequence, syntactic parse tree, dependency parse tree
- A bottom-up strategy
  - Start with unigram features, and gradually increase the size/complexity of the features
  - First search in each subspace, then merge features from different subspaces

Empirical Evaluation

- Data set:
  - ACE (Automatic Content Extraction) 2004
  - 7 types of relations
- Preprocessing
  - Assume entities are correctly identified
  - Brill Tagger
  - Collins Parser
- Learning algorithms
  - Maximum entropy models
  - SVM

Evaluation

- A commonly used setup:
  - Consider all pairs of entities in each single sentence
  - Multi-class classification: # relation types + 1 (no relation between the two entities)
  - 5-fold cross validation
  - Precision (P), Recall (R) and F1 (F)

Increase Feature Complexity

			Uni	+Bi	+Tri	+Prod
ME	Seq	P	0.647	0.662	0.717	N/A
		R	0.614	0.701	0.653	
		F	0.630	0.681	0.683	
	Syn	P	0.651	0.695	0.726	0.702
		R	0.645	0.698	0.688	0.691
		F	0.648	0.697	**0.707**	0.696
	Dep	P	0.647	0.673	0.718	N/A
		R	0.614	0.676	0.652	
		F	0.630	0.674	0.683	
SVM	Seq	P	0.583	0.666	0.684	N/A
		R	0.586	0.650	0.648	
		F	0.585	0.658	0.665	
	Syn	P	0.598	0.645	0.679	0.674
		R	0.611	0.663	0.681	0.672
		F	0.604	0.654	**0.680**	0.673
	Dep	P	0.583	0.644	0.682	N/A
		R	0.586	0.638	0.645	
		F	0.585	0.641	0.663	

Combine Features from Different Subspaces

		Syn	Syn + Seq	Syn + Dep	All
ME	**P**	0.726	0.737	0.695	0.724
	R	0.688	0.694	0.731	0.702
	F	0.683	**0.715**	0.712	0.713
SVM	**P**	0.679	0.689	0.687	0.691
	R	0.681	0.686	0.682	0.686
	F	0.680	**0.688**	0.684	0.688

Heuristics to Prune Features

- H1: in Syn, to remove words before and after the arguments
- H2: in Seq, to remove features that contain articles, adjectives and adverbs
- H3: in Syn, to remove features that contain articles, adjectives and adverbs
- H4: in Seq, to remove words before and after the arguments

Effects of Heuristics

	ME			SVM		
	P	R	F	P	R	F
Best	0.737	0.694	0.715	0.689	0.686	0.688
+H1	0.714	0.729	0.721	0.698	0.699	0.699
+H2	0.730	0.723	0.726	0.704	0.704	**0.704**
+H3	0.739	0.704	0.721	0.701	0.696	0.698
-H3+H4	0.746	0.713	**0.729**	0.702	0.701	0.702

Conclusions

- A general graphic view of feature space
- Evaluated 3 subspaces (seq, syn, dep)
- Findings
  - Combination of unigrams, bigrams, and trigrams works the best
  - Combination of complementary feature subspaces (seq + syn) is beneficial
  - Additional heuristics can be used to further improve the performance

Future Work

- Best feature configuration and relation types
- Principled ways to prune or to weight features
  - Feature selection (information gain, chi square, etc.)
  - Inclusion of more complex features
  - Feature weighting

References

- [Bunescu & Mooney 05a] A shortest path dependency kernel for relation extraction. In *Proceedings of HLT/EMNLP*, 2005.
- [Bunescu & Mooney 05b] Subsequence kernels for relation extraction. In *NIPS*, 2005.
- [Califf & Mooney 98] Relational learning of pattern-match rules for information extraction. In *Proceedings of AAAI Spring Symposium on Applying Machine Learning to Discourse Processing*, 1998.
- [Miller et al. 00] A novel use of statistical parsing to extract information from text. In *Proceedings of NAACL*, 2000.
- [Zhang et al. 06] Exploring syntactic features for relation extraction using a convolution tree kernel. In *Proceedings of HLT/NAACL*, 2006.
- [Zhao & Grishman 05] Extracting relations with integrated information using kenrel methods. In *Proceedings of ACL*, 2005.
- [Zhou et al. 05] Exploring various knowledge in relation extraction. In *Proceedings of ACL*, 2005.

Thanks!