Semantic Class Induction and Coreference Resolution

Vincent Ng
Human Language Technology Research Institute
University of Texas at Dallas

Goal

Improve ACE coreference resolution using automatically acquired semantic class (SC) information of NPs

ACE Coreference

- ACE 2003 Coreference
  - Resolve references to NPs that belong to one of the five semantic classes (entity types)

ACE Coreference

- ACE 2003 Coreference
  - Resolve references to NPs that belong to one of the five semantic classes (entity types)
- PERSON
- ORGANIZATION
- FACILITY
- GPE
- LOCATION

ACE Coreference

- ACE 2003 Coreference
  - Resolve references to NPs that belong to one of the five semantic classes (entity types)
- PERSON (human)
  - George Bush, the postman, …
- ORGANIZATION
- FACILITY
- GPE
- LOCATION

ACE Coreference

- ACE 2003 Coreference
  - Resolve references to NPs that belong to one of the five semantic classes (entity types)
- PERSON (human)
  - George Bush, the postman, …
- ORGANIZATION (corporation, agency, institution)
  - Massachusetts Institute of Technology, the company, …
- FACILITY
- GPE
- LOCATION

ACE Coreference

- ACE 2003 Coreference
  - Resolve references to NPs that belong to one of the five semantic classes (entity types)
- PERSON (human)
  - George Bush, the postman, …
- ORGANIZATION (corporation, agency, institution)
  - Massachusetts Institute of Technology, the company, …
- FACILITY (man-made structure)
  - Sapporo Convention Center, the building, the museum, …
- GPE
- LOCATION

ACE Coreference

- ACE 2003 Coreference
  - Resolve references to NPs that belong to one of the five semantic classes (entity types)
- PERSON (human)
  - George Bush, the postman, …
- ORGANIZATION (corporation, agency, institution)
  - Massachusetts Institute of Technology, the company, …
- FACILITY (man-made structure)
  - Sapporo Convention Center, the building, the museum, …
- GPE (geo-political region)
  - Prague, Czech Republic, the city, the province, …
- LOCATION

ACE Coreference

- ACE 2003 Coreference
  - Resolve references to NPs that belong to one of the five semantic classes (entity types)

- PERSON (human)
  - George Bush, the postman, …
- ORGANIZATION (corporation, agency, institution)
  - Massachusetts Institute of Technology, the company, …
- FACILITY (man-made structure)
  - Sapporo Convention Center, the building, the museum, …
- GPE (geo-political region)
  - Prague, Czech Republic, the city, the province, …
- LOCATION (geographical area, landmass, body of water)
  - River Rhine, the Himalayas, the mountain, …

Using Semantic Class Information for Coreference Resolution

Given the SC of an NP, derive two knowledge sources:

Using Semantic Class Information for Coreference Resolution

Given the SC of an NP, derive two knowledge sources:

1. Semantic class agreement (SCA)
   - yes if and only if both NPs have the same SC

Using Semantic Class Information for Coreference Resolution

Given the SC of an NP, derive two knowledge sources:

1. Semantic class agreement (SCA)
   - yes if and only if both NPs have the same SC
   - Yes for President Bush and the girl

Using Semantic Class Information for Coreference Resolution

Given the SC of an NP, derive two knowledge sources:

1. Semantic class agreement (SCA)
   - yes if and only if both NPs have the same SC
   - Yes for President Bush and the girl
   - No for Prague and the president

Using Semantic Class Information for Coreference Resolution

Given the SC of an NP, derive two knowledge sources:

1. Semantic class agreement (SCA)
   - yes if and only if both NPs have the same SC
   - Yes for President Bush and the girl
   - No for Prague and the president
   - disallow coreference between semantically incompatible NPs

Using Semantic Class Information for Coreference Resolution

Given the SC of an NP, derive two knowledge sources:

1. Semantic class agreement (SCA)
   - yes if and only if both NPs have the same SC
   - Yes for President Bush and the girl
   - No for Prague and the president
   - disallow coreference between semantically incompatible NPs

2. Mention

Using Semantic Class Information for Coreference Resolution

Given the SC of an NP, derive two knowledge sources:

1. Semantic class agreement (SCA)
   - yes if and only if both NPs have the same SC
   - Yes for President Bush and the girl
   - No for Prague and the president
   - disallow coreference between semantically incompatible NPs

2. Mention
   - Defined on an NP
   - yes if and only if the NP belongs to one of the five ACE SCs

Using Semantic Class Information for Coreference Resolution

Given the SC of an NP, derive two knowledge sources:

1. Semantic class agreement (SCA)
   - yes if and only if both NPs have the same SC
   - Yes for President Bush and the girl
   - No for Prague and the president
   - disallow coreference between semantically incompatible NPs

2. Mention
   - Defined on an NP
   - yes if and only if the NP belongs to one of the five ACE SCs
   - Yes for President Bush; No for the dictionary, the satellite …

Using Semantic Class Information for Coreference Resolution

Given the SC of an NP, derive two knowledge sources:

1. Semantic class agreement (SCA)
   - determines whether the SCs of two NPs agree or not
   - Yes for President Bush and the girl
   - No for Prague and the president
   - disallow coreference between semantically incompatible NPs

2. Mention
   - defined on an NP
   - yes if and only if the NP belongs to one of the five ACE SCs
   - Yes for President Bush; No for the dictionary, the satellite …
   - disallow coreference between two NPs if either (or both) of them is not a mention

Using Semantic Class Information for Coreference Resolution

Improve precision

Given the SC of an NP, derive two knowledge sources:

1. Semantic class agreement (SCA)
   - yes if and only if both NPs have the same SC
   - Yes for President Bush and the girl
   - No for Prague and the president
   - disallow coreference between semantically incompatible NPs

2. Mention
   - Defined on an NP
   - yes if and only if the NP belongs to one of the five ACE SCs
   - Yes for President Bush; No for the dictionary, the satellite …
   - disallow coreference between two NPs if either (or both) of them is not a mention

Using Semantic Class Information for Coreference Resolution: System Architectures

- Architecture 1
  1. Extract and classify the mentions simultaneously
  2. Train a coreference model on the mentions
  3. Disallow coreference between semantically incompatible NPs

Using Semantic Class Information for Coreference Resolution: System Architectures

- Architecture 1
  1. Extract and determine the SCs of the mentions simultaneously
  2. Train a coreference model on the mentions
  3. Disallow coreference between semantically incompatible NPs

- Architecture 2
  1. Extract all the NPs
  2. Determine the SC of each extracted NP
  3. Train a coreference model on all the NPs
  4. Disallow coreference between semantically incompatible NPs

Plan for the Talk

- Inducing semantic classes

- Using semantic class information for coreference resolution

Inducing Semantic Classes

Inducing Semantic Classes

- A supervised learning approach
  - Train a six-class classifier to classify an NP as PERSON, ORGANIZATION, GPE, FACILITY, LOCATION or OTHERS

Training Corpus

- BBN Entity Type Corpus (Weischedel and Brunstein, 2005)
  - all the Penn Treebank WSJ articles with the ACE mentions manually identified and annotated with their SCs

Training Instance Creation

- One instance for each automatically extracted NP_i
  - 310K instances created
- Class value derived from the training corpus
  - labeled as OTHERS if the NP is not one of the five SCs
- Represented by seven types of features

1. The WN-CLASS Feature

1. The WN-CLASS Feature

- Identify the WordNet keywords related to the five SCs

Semantic Class	**WordNet Keywords**
PERSON	person
ORGANIZATION	social group
GPE	country, province, government, town, city, administration, society, island, community
FACILITY	establishment, construction, building, facility, workplace
LOCATION	dry land, region, landmass, body of water, geographical area, geological formation

1. The WN-CLASS Feature

- Identify the WordNet keywords related to the five SCs

Semantic Class	**WordNet Keywords**
PERSON	person
ORGANIZATION	social group
GPE	country, province, government, town, city, administration, society, island, community
FACILITY	establishment, construction, building, facility, workplace
LOCATION	dry land, region, landmass, body of water, geographical area, geological formation

1. The WN-CLASS Feature

- Identify the WordNet keywords related to the five SCs

Semantic Class	WordNet Keywords
PERSON	person
ORGANIZATION	social group
GPE	country, province, government, town, city, administration, society, island, community
FACILITY	establishment, construction, building, facility, workplace
LOCATION	dry land, region, landmass, body of water, geographical area, geological formation

- Feature value is the keyword that is a hypernym of the head noun of NP_i

1. The WN-CLASS Feature

- Identify the WordNet keywords related to the five SCs

Semantic Class	WordNet Keywords
PERSON	person
ORGANIZATION	social group
GPE	country, province, government, town, city, administration, society, island, community
FACILITY	establishment, construction, building, facility, workplace
LOCATION	dry land, region, landmass, body of water, geographical area, geological formation

- Feature value is the keyword that is a hypernym of the head noun of NP_i

Bay of Bengal

1. The WN-CLASS Feature

- Identify the WordNet keywords related to the five SCs

Semantic Class	**WordNet Keywords**
PERSON	person
ORGANIZATION	social group
GPE	country, province, government, town, city, administration, society, island, community
FACILITY	establishment, construction, building, facility, workplace
LOCATION	dry land, region, landmass, body of water, geographical area, geological formation

- Feature value is the keyword that is a hypernym of the head noun of NP_i
  - use only first sense to determine if hypernym relation exists

1. The WN-CLASS Feature

- Identify the WordNet keywords related to the five SCs

Semantic Class	WordNet Keywords
PERSON	person
ORGANIZATION	social group
GPE	country, province, government, town, city, administration, society, island, community
FACILITY	establishment, construction, building, facility, workplace
LOCATION	dry land, region, landmass, body of water, geographical area, geological formation

- Feature value is the keyword that is a hypernym of the head noun of NP_i
  - use only first sense to determine if hypernym relation exists
  - no feature is created if no hypernym relation exists

2. The INDUCED-CLASS Feature

- Feature value is the induced class of NP_i

2. The INDUCED-CLASS Feature

- Feature value is the induced class of NP_i

- Given a large, unannotated corpus
  - Extract appositive relations
    - <Eastern Airlines, carrier>, <George Bush, president>, …
  - Use a named entity (NE) recognizer to find the semantic classes of the proper names
  - Infer the semantic class of a common nouns from the associated proper name

2. The INDUCED-CLASS Feature

- Feature value is the induced class of NP_i
- Given a large, unannotated corpus ← **BLLIP+Reuters**
  - Extract appositive relations
    - <Eastern Airlines, carrier>, <George Bush, president>, …
  - Use a named entity (NE) recognizer to find the semantic classes of the proper names
  - Infer the semantic class of a common nouns from the associated proper name

2. The INDUCED-CLASS Feature

- Feature value is the induced class of NP_i

- Given a large, unannotated corpus ← **BLLIP+Reuters**
  - Extract appositive relations ← **MINIPAR**
    - <Eastern Airlines, carrier>, <George Bush, president>, …
  - Use a named entity (NE) recognizer to find the semantic classes of the proper names
  - Infer the semantic class of a common nouns from the associated proper name

2. The INDUCED-CLASS Feature

- Feature value is the induced class of NP_i

- Given a large, unannotated corpus ← **BLLIP+Reuters**
  - Extract appositive relations ← **MINIPAR**
    - <Eastern Airlines, carrier>, <George Bush, president>, …
  - Use a named entity (NE) recognizer to find the semantic classes of the proper names **Identifinder (MUC-style NER)**
  - Infer the semantic class of a common nouns from the associated proper name

Potential Problems

- IdentiFinder is not perfect
  - Mislabels proper names
- MINIPAR is not perfect
  - Extracts NP pairs that are not in apposition

Potential Problems

- IdentiFinder is not perfect
  - Mislabels proper names
- MINIPAR is not perfect
  - Extracts NP pairs that are not in apposition

a common noun could be labeled with more than one semantic class

Potential Problems

- IdentiFinder is not perfect
  - Mislabels proper names
- MINIPAR is not perfect
  - Extracts NP pairs that are not in apposition

a common noun could be labeled with more than one semantic class

- Need a more robust method of inferring the semantic class of a common noun

Potential Problems

- IdentiFinder is not perfect
  - Mislabels proper names
- MINIPAR is not perfect
  - Extracts NP pairs that are not in apposition

a common noun could be labeled with more than one semantic class

- Need a more robust method of inferring the semantic class of a common noun
  1. Compute the probability that the common noun co-occurs with each of the named entity types
  2. If the most likely NE type has a probability above 0.7, label the common noun with the most likely NE type

Other Problems

- Common nouns that do not belong to one of the seven MUC NE types will remain unlabeled

Other Problems

- Common nouns that do not belong to one of the seven MUC NE types will remain unlabeled

- Common nouns that do not co-occur with any NE type with a probability above 0.7 will remain unlabeled

Other Problems

- Common nouns that do not belong to one of the seven MUC NE types will remain unlabeled
- Common nouns that do not co-occur with any NE type with a probability above 0.7 will remain unlabeled
- Solution: fall back on the first-sense heuristic

3. The WORD Feature

- Create one feature for each word in NP_i whose value is the word itself
  - No features are created for stopwords

3. The WORD Feature

- Create one feature for each word in NP_i whose value is the word itself
  - No features are created for stopwords

- Given “the red ball”, we create “red” and “ball” as features

4. The NEIGHBOUR Features

- Feature values are NP_i's semantically similar words

4. The NEIGHBOUR Features

- Feature values are NP_i's semantically similar words
- Adopt a distributional approach

4. The NEIGHBOUR Features

- Feature values are NP_i's semantically similar words

- Adopt a distributional approach
  - Use the similarity values provided by Lin's (1998) dependency-based thesaurus

4. The NEIGHBOUR Features

- Feature values are NP_i's semantically similar words

- Adopt a distributional approach
  - Use the similarity values provided by Lin's (1998) dependency-based thesaurus

- Create one feature for each of the 10 words that are most semantically similar to NP_i

5. The NE Feature

- Feature value is the NE type of NP_i as determined by IdentiFinder

5. The NE Feature

- Feature value is the NE type of NP_i as determined by IdentiFinder
  - If NP_i is determined to be a LOCATION, create an NE feature whose value is GPE
    - most MUC LOCATIONs are ACE GPEs)

6. The SUBJ-VERB Feature

- If NP_i is involved in a subject-verb relation, create a feature whose value is the verb participating in the relation

7. The Verb-Object Feature

- If NP_i is involved in a verb-object relation, create a feature whose value is the verb participating in the relation

Learning Algorithms

- Naïve Bayes
- Multi-class perceptron (Crammer and Singer, 2003)
- 1-nearest neighbour (Daelemans et al.’s TiMBL)
- Decision list (Collins and Singer, 1999)
- Maximum entropy (Och’s YASMET)

Evaluating the Classifiers

- Evaluation corpus: the ACE 2003 coreference corpus
  - comprises a training set and a test set

Evaluating the Classifiers

- Evaluation corpus: the ACE 2003 coreference corpus
  - comprises a training set and a test set

- Test instances created in the same way as training instances
  - one instance for each automatically extracted NP
  - training set: 54.6K instances; test set: 13.4K instances

Evaluating the Classifiers

- Evaluation corpus: the ACE 2003 coreference corpus
  - comprises a training set and a test set

- Test instances created in the same way as training instances
  - one instance for each automatically extracted NP
  - training set: 54.6K instances; test set: 13.4K instances

- Class distribution (in percentages):

	PER	ORG	GPE	FAC	LOC	OTH
ACE Training Set (54.6K)	19.8	9.6	11.4	1.6	1.2	56.3
ACE Test Set (13.4K)	19.5	9.0	9.6	1.8	1.1	59.0

Evaluating the Classifiers

- Evaluation corpus: the ACE 2003 coreference corpus
  - comprises a training set and a test set
- Test instances created in the same way as training instances
  - one instance for each automatically extracted NP
  - training set: 54.6K instances; test set: 13.4K instances
- Class distribution (in percentages):

	PER	ORG	GPE	FAC	LOC	OTH
ACE Training Set (54.6K)	19.8	9.6	11.4	1.6	1.2	56.3
ACE Test Set (13.4K)	19.5	9.0	9.6	1.8	1.1	59.0

Evaluating the Classifiers

- Evaluation corpus: the ACE 2003 coreference corpus
  - comprises a training set and a test set
- Test instances created in the same way as training instances
  - one instance for each automatically extracted NP
  - training set: 54.6K instances; test set: 13.4K instances
- Class distribution (in percentages):

	PER	ORG	GPE	FAC	LOC	OTH
ACE Training Set (54.6K)	19.8	9.6	11.4	1.6	1.2	56.3
ACE Test Set (13.4K)	19.5	9.0	9.6	1.8	1.1	59.0

Evaluating the Classifiers

- Evaluation corpus: the ACE 2003 coreference corpus
  - comprises a training set and a test set
- Test instances created in the same way as training instances
  - one instance for each automatically extracted NP
  - training set: 54.6K instances; test set: 13.4K instances
- Class distribution (in percentages):

	PER	ORG	GPE	FAC	LOC	OTH
ACE Training Set (54.6K)	19.8	9.6	11.4	1.6	1.2	56.3
ACE Test Set (13.4K)	19.5	9.0	9.6	1.8	1.1	59.0

Evaluating the Classifiers

- Evaluation corpus: the ACE 2003 coreference corpus
  - comprises a training set and a test set
- Test instances created in the same way as training instances
  - one instance for each automatically extracted NP
  - training set: 54.6K instances; test set: 13.4K instances
- Class distribution (in percentages):

	PER	ORG	GPE	FAC	LOC	OTH
ACE Training Set (54.6K)	19.8	9.6	11.4	1.6	1.2	56.3
ACE Test Set (13.4K)	19.5	9.0	9.6	1.8	1.1	59.0

Evaluating the Classifiers

- Evaluation corpus: the ACE 2003 coreference corpus
  - comprises a training set and a test set
- Test instances created in the same way as training instances
  - one instance for each automatically extracted NP
  - training set: 54.6K instances; test set: 13.4K instances
- Class distribution (in percentages):

	PER	ORG	GPE	FAC	LOC	OTH
ACE Training Set (54.6K)	19.8	9.6	11.4	1.6	1.2	56.3
ACE Test Set (13.4K)	19.5	9.0	9.6	1.8	1.1	59.0

Evaluating the Classifiers

- Evaluation corpus: the ACE 2003 coreference corpus
  - comprises a training set and a test set
- Test instances created in the same way as training instances
  - one instance for each automatically extracted NP
  - training set: 54.6K instances; test set: 13.4K instances
- Class distribution (in percentages):

	PER	ORG	GPE	FAC	LOC	OTH
ACE Training Set (54.6K)	19.8	9.6	11.4	1.6	1.2	56.3
ACE Test Set (13.4K)	19.5	9.0	9.6	1.8	1.1	59.0

Baseline SC Classification Method

- Motivated by Soon et al. (2001)

Baseline SC Classification Method

- Motivated by Soon et al. (2001)
  - Assign to a named entity its NE feature value

Baseline SC Classification Method

- Motivated by Soon et al. (2001)
  - Assign to a named entity its NE feature value
  - Assign to any other NP the semantic class associated with its WN-CLASS feature value

Baseline SC Classification Method

- Motivated by Soon et al. (2001)
  - Assign to a named entity its NE feature value
  - Assign to any other NP the semantic class associated with its WN-CLASS feature value
    - "Bay of Bengal" "body of water" LOCATION

Classification Accuracies

	ACE Training Set		
	Proper	Common	Overall
Baseline	83.1	83.1	**83.1**
Naïve Bayes	65.5	73.8	**71.3**
Perceptron	75.7	86.4	**83.2**
1-NN	81.0	85.2	**84.0**
Decision list	84.1	85.4	**85.0**
MaxEnt	80.9	87.0	**85.2**

Classification Accuracies

	ACE Training Set		
	Proper	Common	Overall
Baseline	83.1	83.1	**83.1**
Naïve Bayes	65.5	73.8	**71.3**
Perceptron	75.7	86.4	**83.2**
1-NN	81.0	85.2	**84.0**
Decision list	84.1	85.4	**85.0**
MaxEnt	80.9	87.0	**85.2**

Classification Accuracies

	ACE Training Set		
	Proper	Common	Overall
Baseline	83.1	83.1	**83.1**
Naïve Bayes	65.5	73.8	**71.3**
Perceptron	75.7	86.4	**83.2**
1-NN	81.0	85.2	**84.0**
Decision list	84.1	85.4	**85.0**
MaxEnt	80.9	87.0	**85.2**

Classification Accuracies

	ACE Training Set		
	Proper	Common	Overall
Baseline	83.1	83.1	**83.1**
Naïve Bayes	65.5	73.8	**71.3**
Perceptron	75.7	86.4	**83.2**
1-NN	81.0	85.2	**84.0**
Decision list	84.1	85.4	**85.0**
MaxEnt	80.9	87.0	**85.2**

Classification Accuracies

	ACE Training Set		
	Proper	Common	Overall
Baseline	83.1	83.1	**83.1**
Naïve Bayes	65.5	73.8	**71.3**
Perceptron	75.7	86.4	**83.2**
1-NN	81.0	85.2	**84.0**
Decision list	84.1	85.4	**85.0**
MaxEnt	80.9	87.0	**85.2**

Classification Accuracies

	ACE Training Set		
	Proper	Common	Overall
Baseline	83.1	83.1	**83.1**
Naïve Bayes	65.5	73.8	**71.3**
Perceptron	75.7	86.4	**83.2**
1-NN	81.0	85.2	**84.0**
Decision list	84.1	85.4	**85.0**
MaxEnt	80.9	87.0	**85.2**

- 1-NN, Decision List, and MaxEnt outperform the baseline significantly

Classification Accuracies

	ACE Training Set		
	Proper	Common	Overall
Baseline	83.1	83.1	**83.1**
Naïve Bayes	65.5	73.8	**71.3**
Perceptron	75.7	86.4	**83.2**
1-NN	81.0	85.2	**84.0**
Decision list	84.1	85.4	**85.0**
MaxEnt	80.9	87.0	**85.2**

- Decision List achieves the highest accuracy on proper NPs

Classification Accuracies

	ACE Training Set		
	Proper	Common	Overall
Baseline	83.1	83.1	**83.1**
Naïve Bayes	65.5	73.8	**71.3**
Perceptron	75.7	86.4	**83.2**
1-NN	81.0	85.2	**84.0**
Decision list	84.1	85.4	**85.0**
MaxEnt	80.9	87.0	**85.2**

- MaxEnt achieves the highest accuracy on proper NPs

Classification Accuracies

	ACE Training Set		
	Proper	Common	Overall
Baseline	83.1	83.1	**83.1**
Naïve Bayes	65.5	73.8	**71.3**
Perceptron	75.7	86.4	**83.2**
1-NN	81.0	85.2	**84.0**
Decision list	84.1	85.4	**85.0**
MaxEnt	80.9	87.0	**85.2**
DLME	84.1	87.0	**86.1**

- DLME: combines the output for Decision List and MaxEnt
  - Uses Decision List for proper NP classification
  - Uses MaxEnt for common NP classification

Classification Accuracies

	ACE Training Set			ACE Test Set		
	Proper	Common	Overall	Proper	Common	Overall
Baseline	83.1	83.1	**83.1**	79.6	81.6	**81.1**
Naïve Bayes	65.5	73.8	**71.3**	64.4	72.6	**70.3**
Perceptron	75.7	86.4	**83.2**	73.4	84.1	**81.2**
1-NN	81.0	85.2	**84.0**	79.8	84.3	**83.1**
Decision list	84.1	85.4	**85.0**	82.0	83.3	**82.9**
MaxEnt	80.9	87.0	**85.2**	78.9	85.7	**83.8**
DLME	84.1	87.0	**86.1**	82.0	85.7	**84.7**

- Results for the test set exhibit similar trends

Classification Accuracies

	ACE Training Set			ACE Test Set		
	Proper	Common	Overall	Proper	Common	Overall
Baseline	83.1	83.1	**83.1**	79.6	81.6	**81.1**
Naïve Bayes	65.5	73.8	**71.3**	64.4	72.6	**70.3**
Perceptron	75.7	86.4	**83.2**	73.4	84.1	**81.2**
1-NN	81.0	85.2	**84.0**	79.8	84.3	**83.1**
Decision list	84.1	85.4	**85.0**	82.0	83.3	**82.9**
MaxEnt	80.9	87.0	**85.2**	78.9	85.7	**83.8**
DLME	84.1	87.0	**86.1**	82.0	85.7	**84.7**

- Results for the test set exhibit similar trends
  - Decision List achieves the best accuracy on proper NPs
  - MaxEnt achieves the best accuracy on common NPs

Classification Accuracies

	ACE Training Set			ACE Test Set		
	Proper	Common	Overall	Proper	Common	Overall
Baseline	83.1	83.1	**83.1**	79.6	81.6	**81.1**
Naïve Bayes	65.5	73.8	**71.3**	64.4	72.6	**70.3**
Perceptron	75.7	86.4	**83.2**	73.4	84.1	**81.2**
1-NN	81.0	85.2	**84.0**	79.8	84.3	**83.1**
Decision list	84.1	85.4	**85.0**	82.0	83.3	**82.9**
MaxEnt	80.9	87.0	**85.2**	78.9	85.7	**83.8**
DLME	84.1	87.0	**86.1**	82.0	85.7	**84.7**

- DLME has the highest overall accuracy for the test set

Classification Accuracies

	ACE Training Set			ACE Test Set		
	Proper	Common	Overall	Proper	Common	Overall
Baseline	83.1	83.1	**83.1**	79.6	81.6	**81.1**
Naïve Bayes	65.5	73.8	**71.3**	64.4	72.6	**70.3**
Perceptron	75.7	86.4	**83.2**	73.4	84.1	**81.2**
1-NN	81.0	85.2	**84.0**	79.8	84.3	**83.1**
Decision list	84.1	85.4	**85.0**	82.0	83.3	**82.9**
MaxEnt	80.9	87.0	**85.2**	78.9	85.7	**83.8**
DLME	84.1	87.0	**86.1**	82.0	85.7	**84.7**

- DLME has the highest overall accuracy for the test set

1. Which feature types are important for achieving this accuracy?

Classification Accuracies

	ACE Training Set			ACE Test Set		
	Proper	Common	Overall	Proper	Common	Overall
Baseline	83.1	83.1	**83.1**	79.6	81.6	**81.1**
Naïve Bayes	65.5	73.8	**71.3**	64.4	72.6	**70.3**
Perceptron	75.7	86.4	**83.2**	73.4	84.1	**81.2**
1-NN	81.0	85.2	**84.0**	79.8	84.3	**83.1**
Decision list	84.1	85.4	**85.0**	82.0	83.3	**82.9**
MaxEnt	80.9	87.0	**85.2**	78.9	85.7	**83.8**
DLME	84.1	87.0	**86.1**	82.0	85.7	**84.7**

- DLME has the highest overall accuracy for the test set

1. Which feature types are important for achieving this accuracy?
2. Will this accuracy be sufficient for improving a coreference system?

Feature Contribution

- Feature ablation experiments
  - Train classifiers with all but one type of features

Feature Contribution

- Feature ablation experiments
  - Train classifiers with all but one type of features

- Key observations
  - Accuracy for proper NPs drops significantly when the NE features are left out
  - Accuracy for proper NPs drops significantly (but to a lesser extent) when the NEIGHBOUR features are left out
  - Accuracy for common NPs drops moderately (but not significantly) when the INDUCED-CLASS feature is left out

Plan for the Talk

- Inducing semantic classes

- Using semantic class information for coreference resolution

Coreference Resolution:
Standard Machine Learning Approach

Step 1: Classification

- classifies two NPs as *coreferent* or *not coreferent*

Step 2: Clustering

- coordinates pairwise classification decisions
- single-link clustering to find an antecedent for each NP

Incorporating Semantic Class Information

- Two knowledge sources
  - semantic class agreement (SCA): do two NPs agree in SC?
  - mention: does the NP belong to one of the five ACE SCs?

Incorporating Semantic Class Information

- Two knowledge sources
  - semantic class agreement (SCA): do two NPs agree in SC?
  - mention: does the NP belong to one of the five ACE SCs?

- Incorporate the two knowledge sources as
  - Features
  - Constraints

Incorporating Semantic Class Information

- Two knowledge sources
  - semantic class agreement (SCA): do two NPs agree in SC?
  - mention: does the NP belong to one of the five ACE SCs?

- Incorporate the two knowledge sources as
  - Features
  - Constraints
    - SCA: disallow coreference if the NPs disagree w.r.t. SC
    - Mention: disallow coreference if at least one NP is not a mention

Incorporating Semantic Class Information

- Two knowledge sources
  - semantic class agreement (SCA): do two NPs agree in SC?
  - mention: does the NP belong to one of the five ACE SCs?

- Incorporate the two knowledge sources as
  - Features → affects classification
  - Constraints → affects clustering
    - SCA: disallow coreference if the NPs disagree w.r.t. SC
    - Mention: disallow coreference if at least one NP is not a mention

Incorporating Semantic Class Information

- Eight ways of incorporating the two knowledge sources
  - they can be applied in isolation or in combination
  - each can be used as a constraint or as a feature

Incorporating Semantic Class Information

- Eight ways of incorporating the two knowledge sources
  - they can be applied in isolation or in combination
  - each can be used as a constraint or as a feature
- The eight combinations:
  1. Mention (C)
  2. Mention (F)
  3. SCA (C)
  4. SCA (F)
  5. Mention (C) + SCA (C)
  6. Mention (C) + SCA (F)
  7. Mention (F) + SCA (C)
  8. Mention (F) + SCA (F)

Incorporating Semantic Class Information

- Eight ways of incorporating the two knowledge sources
  - they can be applied in isolation or in combination
  - each can be used as a constraint or as a feature

- The eight combinations:

1. Mention (C)
2. Mention (F)
3. SCA (C)
4. SCA (F)
5. Mention (C) + SCA (C)
6. Mention (C) + SCA (F)
7. Mention (F) + SCA (C)
8. Mention (F) + SCA (F)

Incorporating Semantic Class Information

- Eight ways of incorporating the two knowledge sources
  - they can be applied in isolation or in combination
  - each can be used as a constraint or as a feature

- The eight combinations:
  1. Mention (C)
  2. Mention (F)
  3. SCA (C)
  4. SCA (F)
  5. Mention (C) + SCA (C)
  6. Mention (C) + SCA (F)
  7. Mention (F) + SCA (C)
  8. Mention (F) + SCA (F)

Incorporating Semantic Class Information

- Eight ways of incorporating the two knowledge sources
  - they can be applied in isolation or in combination
  - each can be used as a constraint or as a feature

- The eight combinations:
  1. Mention (C)
  2. Mention (F)
  3. SCA (C)
  4. SCA (F)
  5. Mention (C) + SCA (C)
  6. Mention (C) + SCA (F)
  7. Mention (F) + SCA (C)
  8. Mention (F) + SCA (F)

Experimental Setup

- The 2003 ACE coreference corpus
  - train on training set and evaluate on test set
  - NPs extracted automatically

Experimental Setup

- The 2003 ACE coreference corpus
  - train on training set and evaluate on test set
  - NPs extracted automatically
- Evaluation metrics
  - F-measure
    - computed by MUC scoring program (Vilain et al., 1995)

Experimental Setup

- The 2003 ACE coreference corpus
  - train on training set and evaluate on test set
  - NPs extracted automatically
- Evaluation metrics
  - F-measure
    - computed by MUC scoring program (Vilain et al., 1995)
  - accuracy on resolving anaphoric NPs
    - consider an NP correctly resolved if it appears in the same cluster as its closest antecedent (Ponzetto and Strube, 2006)

The Baseline Coreference System

- Learning algorithm: C4.5
- Clustering: single-link clustering
- Training instance creation method: Soon et al. (2001)
- Feature set (33 features):
  - String-matching features
    - Exact string match, substring match, head noun match
  - Grammatical features
    - Agreement w.r.t. gender, number, animacy, grammatical role
  - Positional feature
    - Distance between the two NPs in sentences
  - Semantic feature: Name alias

Results (Baseline System)

	MUC Scorer		
	R	P	F
Baseline system	60.9	53.6	**57.0**

Results (Baseline System)

	MUC Scorer		
	R	P	F
Baseline system	60.9	53.6	**57.0**

Results (Baseline System)

	MUC Scorer		
	R	P	F
Baseline system	60.9	53.6	**57.0**

Results (Baseline System)

	MUC Scorer		
	R	P	F
Baseline system	60.9	53.6	**57.0**

Results (Baseline System)

	MUC Scorer		
	R	P	F
Baseline system	60.9	53.6	**57.0**

- How strong is the baseline result?

Results (Baseline System)

	MUC Scorer		
	R	P	F
Baseline system	60.9	53.6	**57.0**

- How strong is the baseline result?
  - Compare the baseline with the Soon et al. (2001) coreference system

Results (Duplicated Soon et al. System)

	MUC Scorer		
	R	P	F
Baseline system	60.9	53.6	**57.0**
Duplicated Soon et al.	56.1	54.4	**55.3**

- How strong is the baseline result?
  - Compare the baseline with the Soon et al. (2001) coreference system

Results (Duplicated Soon et al. System)

	MUC Scorer		
	R	P	F
Baseline system	60.9	53.6	**57.0**
Duplicated Soon et al.	56.1	54.4	**55.3**

- How strong is the baseline result?
  - Compare the baseline with the Soon et al. (2001) coreference system
  - Performance difference is highly significant (*p*=0.002)

Incorporating Semantic Class Information

- The SC of a proper or common NP is given by the DLME classifier.

Incorporating Semantic Class Information

- The SC of a proper or common NP is given by the DLME classifier.

- The SC of a pronoun is UNCONSTRAINED (i.e., it is semantically compatible with all other NPs).

Incorporating Semantic Class Information

- The SC of a proper or common NP is given by the DLME classifier.
- The SC of a pronoun is UNCONSTRAINED (i.e., it is semantically compatible with all other NPs).
- Derive SCA and Mention from the induced SCs.
- Incorporate knowledge sources into coreference system.

Results (Using Induced SC Information)

	MUC Scorer		
	R	P	F
Baseline system	60.9	53.6	**57.0**
Add to the Baseline system			
Mention (C) only	58.7	72.0	**64.7**
Mention (F) only	61.3	53.7	**57.3**
SCA (C) only	57.3	72.0	**63.8**
SCA (F) only	62.9	54.9	**58.6**
Mention (C) + SCA (C)	57.5	72.2	**64.0**
Mention (C) + SCA (F)	61.0	69.6	**65.0**
Mention (F) + SCA (C)	57.6	72.2	**64.1**
Mention (F) + SCA (F)	63.2	53.4	**57.9**

Results (Using Induced SC Information)

	MUC Scorer		
	R	P	F
Baseline system	60.9	53.6	**57.0**
Add to the Baseline system			
Mention (C) only	58.7	72.0	**64.7**
Mention (F) only	61.3	53.7	**57.3**
SCA (C) only	57.3	72.0	**63.8**
SCA (F) only	62.9	54.9	**58.6**
Mention (C) + SCA (C)	57.5	72.2	**64.0**
Mention (C) + SCA (F)	61.0	69.6	**65.0**
Mention (F) + SCA (C)	57.6	72.2	**64.1**
Mention (F) + SCA (F)	63.2	53.4	**57.9**

- Significant improvements over the baseline in six cases

Results (Using Induced SC Information)

	MUC Scorer		
	R	P	F
Baseline system	60.9	53.6	**57.0**
Add to the Baseline system			
Mention (C) only	58.7	72.0	**64.7**
Mention (F) only	61.3	53.7	**57.3**
SCA (C) only	57.3	72.0	**63.8**
SCA (F) only	62.9	54.9	**58.6**
Mention (C) + SCA (C)	57.5	72.2	**64.0**
Mention (C) + SCA (F)	61.0	69.6	**65.0**
Mention (F) + SCA (C)	57.6	72.2	**64.1**
Mention (F) + SCA (F)	63.2	53.4	**57.9**

- In five cases, F-measure increases by about 7-8
  - large gains in precision and smaller loss in recall

Results (Using Induced SC Information)

	MUC Scorer		
	R	P	F
Baseline system	60.9	53.6	**57.0**
Add to the Baseline system			
Mention (C) only	58.7	72.0	**64.7**
Mention (F) only	61.3	53.7	**57.3**
SCA (C) only	57.3	72.0	**63.8**
SCA (F) only	62.9	54.9	**58.6**
Mention (C) + SCA (C)	57.5	72.2	**64.0**
Mention (C) + SCA (F)	61.0	69.6	**65.0**
Mention (F) + SCA (C)	57.6	72.2	**64.1**
Mention (F) + SCA (F)	63.2	53.4	**57.9**

- At least one of the two knowledge sources are used as constraints

Results (Using Induced SC Information)

	MUC Scorer R	P	F
Baseline system	60.9	53.6	**57.0**
Add to the Baseline system			
Mention (C) only	58.7	72.0	**64.7**
Mention (F) only	61.3	53.7	**57.3**
SCA (C) only	57.3	72.0	**63.8**
SCA (F) only	62.9	54.9	**58.6**
Mention (C) + SCA (C)	57.5	72.2	**64.0**
Mention (C) + SCA (F)	61.0	69.6	**65.0**
Mention (F) + SCA (C)	57.6	72.2	**64.1**
Mention (F) + SCA (F)	63.2	53.4	**57.9**

- At least one of the two knowledge sources are used as constraints

Results (Using Induced SC Information)

	MUC Scorer		
	R	P	F
Baseline system	60.9	53.6	**57.0**
Add to the Baseline system			
Mention (C) only	58.7	72.0	**64.7**
Mention (F) only	61.3	53.7	**57.3**
SCA (C) only	57.3	72.0	**63.8**
SCA (F) only	62.9	54.9	**58.6**
Mention (C) + SCA (C)	57.5	72.2	**64.0**
Mention (C) + SCA (F)	61.0	69.6	**65.0**
Mention (F) + SCA (C)	57.6	72.2	**64.1**
Mention (F) + SCA (F)	63.2	53.4	**57.9**

- At least one of the two knowledge sources are used as constraints

Results (Using Induced SC Information)

	MUC Scorer R	MUC Scorer P	MUC Scorer F
Baseline system	60.9	53.6	**57.0**
Add to the Baseline system			
Mention (C) only	58.7	72.0	**64.7**
Mention (F) only	61.3	53.7	**57.3**
SCA (C) only	57.3	72.0	**63.8**
SCA (F) only	62.9	54.9	**58.6**
Mention (C) + SCA (C)	57.5	72.2	**64.0**
Mention (C) + SCA (F)	61.0	69.6	**65.0**
Mention (F) + SCA (C)	57.6	72.2	**64.1**
Mention (F) + SCA (F)	63.2	53.4	**57.9**

- At least one of the two knowledge sources are used as constraints

Results (Using Induced SC Information)

	MUC Scorer R	P	F
Baseline system	60.9	53.6	**57.0**
Add to the Baseline system			
Mention (C) only	58.7	72.0	**64.7**
Mention (F) only	61.3	53.7	**57.3**
SCA (C) only	57.3	72.0	**63.8**
SCA (F) only	62.9	54.9	**58.6**
Mention (C) + SCA (C)	57.5	72.2	**64.0**
Mention (C) + SCA (F)	61.0	69.6	**65.0**
Mention (F) + SCA (C)	57.6	72.2	**64.1**
Mention (F) + SCA (F)	63.2	53.4	**57.9**

- At least one of the two knowledge sources are used as constraints

Is SCA useful in the presence of Mention (C)?

Results (Using Induced SC Information)

	MUC Scorer			Resolution Accuracy			
	R	P	F	PRO	PN	CN	All
Baseline system	60.9	53.6	**57.0**	59.2	54.8	22.5	**48.4**
Add to the Baseline system							
Mention (C) only	58.7	72.0	**64.7**	58.9	53.3	19.1	**46.8**
Mention (F) only	61.3	53.7	**57.3**	59.2	55.7	22.7	**48.8**
SCA (C) only	57.3	72.0	**63.8**	57.8	51.0	17.0	**45.1**
SCA (F) only	62.9	54.9	**58.6**	59.4	57.1	29.9	**51.2**
Mention (C) + SCA (C)	57.5	72.2	**64.0**	57.9	51.2	17.1	**45.2**
Mention (C) + SCA (F)	61.0	69.6	**65.0**	59.4	56.3	27.2	**50.2**
Mention (F) + SCA (C)	57.6	72.2	**64.1**	57.9	51.5	17.1	**45.4**
Mention (F) + SCA (F)	63.2	53.4	**57.9**	59.7	57.7	30.1	**51.5**

Results (Using Induced SC Information)

	MUC Scorer			Resolution Accuracy			
	R	P	F	PRO	PN	CN	All
Baseline system	60.9	53.6	**57.0**	59.2	54.8	22.5	**48.4**
Add to the Baseline system							
Mention (C) only	58.7	72.0	**64.7**	58.9	53.3	19.1	**46.8**
Mention (F) only	61.3	53.7	**57.3**	59.2	55.7	22.7	**48.8**
SCA (C) only	57.3	72.0	**63.8**	57.8	51.0	17.0	**45.1**
SCA (F) only	62.9	54.9	**58.6**	59.4	57.1	29.9	**51.2**
Mention (C) + SCA (C)	57.5	72.2	**64.0**	57.9	51.2	17.1	**45.2**
Mention (C) + SCA (F)	61.0	69.6	**65.0**	59.4	56.3	27.2	**50.2**
Mention (F) + SCA (C)	57.6	72.2	**64.1**	57.9	51.5	17.1	**45.4**
Mention (F) + SCA (F)	63.2	53.4	**57.9**	59.7	57.7	30.1	**51.5**

Results (Using Induced SC Information)

	MUC Scorer			Resolution Accuracy			
	R	P	F	PRO	PN	CN	All
Baseline system	60.9	53.6	**57.0**	59.2	54.8	22.5	**48.4**
Add to the Baseline system							
Mention (C) only	58.7	72.0	**64.7**	58.9	53.3	19.1	**46.8**
Mention (F) only	61.3	53.7	**57.3**	59.2	55.7	22.7	**48.8**
SCA (C) only	57.3	72.0	**63.8**	57.8	51.0	17.0	**45.1**
SCA (F) only	62.9	54.9	**58.6**	59.4	57.1	29.9	**51.2**
Mention (C) + SCA (C)	57.5	72.2	**64.0**	57.9	51.2	17.1	**45.2**
Mention (C) + SCA (F)	61.0	69.6	**65.0**	59.4	56.3	27.2	**50.2**
Mention (F) + SCA (C)	57.6	72.2	**64.1**	57.9	51.5	17.1	**45.4**
Mention (F) + SCA (F)	63.2	53.4	**57.9**	59.7	57.7	30.1	**51.5**

Results (Using Induced SC Information)

	MUC Scorer			Resolution Accuracy			
	R	P	F	PRO	PN	CN	All
Baseline system	60.9	53.6	**57.0**	59.2	54.8	22.5	**48.4**
Add to the Baseline system							
Mention (C) only	58.7	72.0	**64.7**	58.9	53.3	19.1	**46.8**
Mention (F) only	61.3	53.7	**57.3**	59.2	55.7	22.7	**48.8**
SCA (C) only	57.3	72.0	**63.8**	57.8	51.0	17.0	**45.1**
SCA (F) only	62.9	54.9	**58.6**	59.4	57.1	29.9	**51.2**
Mention (C) + SCA (C)	57.5	72.2	**64.0**	57.9	51.2	17.1	**45.2**
Mention (C) + SCA (F)	61.0	69.6	**65.0**	59.4	56.3	27.2	**50.2**
Mention (F) + SCA (C)	57.6	72.2	**64.1**	57.9	51.5	17.1	**45.4**
Mention (F) + SCA (F)	63.2	53.4	**57.9**	59.7	57.7	30.1	**51.5**

Results (Using Induced SC Information)

	MUC Scorer			Resolution Accuracy			
	R	P	F	PRO	PN	CN	All
Baseline system	60.9	53.6	**57.0**	59.2	54.8	22.5	**48.4**
Add to the Baseline system							
Mention (C) only	58.7	72.0	**64.7**	58.9	53.3	19.1	**46.8**
Mention (F) only	61.3	53.7	**57.3**	59.2	55.7	22.7	**48.8**
SCA (C) only	57.3	72.0	**63.8**	57.8	51.0	17.0	**45.1**
SCA (F) only	62.9	54.9	**58.6**	59.4	57.1	29.9	**51.2**
Mention (C) + SCA (C)	57.5	72.2	**64.0**	57.9	51.2	17.1	**45.2**
Mention (C) + SCA (F)	61.0	69.6	**65.0**	59.4	56.3	27.2	**50.2**
Mention (F) + SCA (C)	57.6	72.2	**64.1**	57.9	51.5	17.1	**45.4**
Mention (F) + SCA (F)	63.2	53.4	**57.9**	59.7	57.7	30.1	**51.5**

Results (Using Induced SC Information)

	MUC Scorer			Resolution Accuracy			
	R	P	F	PRO	PN	CN	All
Baseline system	60.9	53.6	**57.0**	59.2	54.8	22.5	**48.4**
Add to the Baseline system							
Mention (C) only	58.7	72.0	**64.7**	58.9	53.3	19.1	**46.8**
Mention (F) only	61.3	53.7	**57.3**	59.2	55.7	22.7	**48.8**
SCA (C) only	57.3	72.0	**63.8**	57.8	51.0	17.0	**45.1**
SCA (F) only	62.9	54.9	**58.6**	59.4	57.1	29.9	**51.2**
Mention (C) + SCA (C)	57.5	72.2	**64.0**	57.9	51.2	17.1	**45.2**
Mention (C) + SCA (F)	61.0	69.6	**65.0**	59.4	56.3	27.2	**50.2**
Mention (F) + SCA (C)	57.6	72.2	**64.1**	57.9	51.5	17.1	**45.4**
Mention (F) + SCA (F)	63.2	53.4	**57.9**	59.7	57.7	30.1	**51.5**

Results (Using Induced SC Information)

	MUC Scorer			Resolution Accuracy			
	R	P	F	PRO	PN	CN	All
Baseline system	60.9	53.6	**57.0**	59.2	54.8	22.5	**48.4**
Add to the Baseline system							
Mention (C) only	58.7	72.0	**64.7**	58.9	53.3	19.1	**46.8**
Mention (F) only	61.3	53.7	**57.3**	59.2	55.7	22.7	**48.8**
SCA (C) only	57.3	72.0	**63.8**	57.8	51.0	17.0	**45.1**
SCA (F) only	62.9	54.9	**58.6**	59.4	57.1	29.9	**51.2**
Mention (C) + SCA (C)	57.5	72.2	**64.0**	57.9	51.2	17.1	**45.2**
Mention (C) + SCA (F)	61.0	69.6	**65.0**	59.4	56.3	27.2	**50.2**
Mention (F) + SCA (C)	57.6	72.2	**64.1**	57.9	51.5	17.1	**45.4**
Mention (F) + SCA (F)	63.2	53.4	**57.9**	59.7	57.7	30.1	**51.5**

- Mention (C) + SCA (F) is better in terms of overall accuracy

Results (Using Induced SC Information)

	MUC Scorer			Resolution Accuracy			
	R	P	F	PRO	PN	CN	All
Baseline system	60.9	53.6	**57.0**	59.2	54.8	22.5	**48.4**
Add to the Baseline system							
Mention (C) only	58.7	72.0	**64.7**	58.9	53.3	19.1	**46.8**
Mention (F) only	61.3	53.7	**57.3**	59.2	55.7	22.7	**48.8**
SCA (C) only	57.3	72.0	**63.8**	57.8	51.0	17.0	**45.1**
SCA (F) only	62.9	54.9	**58.6**	59.4	57.1	29.9	**51.2**
Mention (C) + SCA (C)	57.5	72.2	**64.0**	57.9	51.2	17.1	**45.2**
Mention (C) + SCA (F)	61.0	69.6	**65.0**	59.4	56.3	27.2	**50.2**
Mention (F) + SCA (C)	57.6	72.2	**64.1**	57.9	51.5	17.1	**45.4**
Mention (F) + SCA (F)	63.2	53.4	**57.9**	59.7	57.7	30.1	**51.5**

- Mention (C) + SCA (F) is better in terms of overall accuracy
  - Outperforms Mention (C) by 3% in proper NP resolution and 8% in common NP resolution

Summary

- Mention and SCA can be usefully employed to improve the performance of a learning-based coreference system

- Experimental results suggest that Mention should be used as a constraint and SCA as a feature.