Coreference Resolution

TALP Research Center

Introduction

The goal of coreference resolution

Determining which mentions in a discourse refer to the same real world entity, property or situation.

Example:

FC Barcelona president Joan Laporta has warned Chelsea off star strike Lionel Messi.

This warning has generated dicouragement in Chelsea.

Aware of Chelsea owner Roman Abramovich's interest in the young Argentine, Laporta said last night: " I will answer as always, Messi is not for sale and we do not want to let him go."

The goal of coreference resolution

Determining which mentions in a discourse refer to the same real-world entity, property or situation.

Example:

FC Barcelona president Joan Laporta has warned Chelsea off star strike Lionel Messi.

This warning has generated dicouragement in Chelsea.

Aware of Chelsea owner Roman Abramovich's interest in the young Argentine, Laporta said last night: " I will answer as always, Messi is not for sale and we do not want to let him go."

The goal of coreference resolution

Determining which mentions in a discourse refer to the same real world entity, property or situation.

Example:

FC Barcelona president Joan Laporta has warned Chelsea off star strike Lionel Messi.

This warning has generated dicouragement in Chelsea.

Aware of Chelsea owner Roman Abramovich's interest in the young Argentine, Laporta said last night: " I will answer as always, Messi is not for sale and we do not want to let him go."

Introduction

General Goal

Types of coreference

Identity noun phrase coreference

Types of coreference (positional viewpoint)

Given two mentions,

- **Anaphora (endophora)**:
  - [*Messi*]$_1$ is not for sale. We do not want to let [*him*]$_1$ go.
  - [*Laporta warned Chelsea off Messi*]$_1$. [*This warning*]$_1$ generated discouragement in Chelsea.
  - [*The car*]$_1$ hit a tree. [*The vehicle*]$_1$ was found one day later.

Types of coreference (positional viewpoint)

Given two mentions,

- **Anaphora (endophora)**:
  - [*Messi*]$_1$ is not for sale. We do not want to let [*him*]$_1$ go.
  - [*Laporta warned Chelsea off Messi*]$_1$. [*This warning*]$_1$ generated discouragement in Chelsea.
  - [*The car*]$_1$ hit a tree. [*The vehicle*]$_1$ was found one day later.
- **Cataphora (endophora)**:
  - We do not want to let [*him*]$_1$ go. [*Messi*]$_1$ is not for sale.

Types of coreference (positional viewpoint)

Given two mentions,

- **Anaphora (endophora)**:
  - [*Messi*]$_1$ is not for sale. We do not want to let [*him*]$_1$ go.
  - [*Laporta warned Chelsea off Messi*]$_1$. [*This warning*]$_1$ generated discouragement in Chelsea.
  - [*The car*]$_1$ hit a tree. [*The vehicle*]$_1$ was found one day later.
- **Cataphora (endophora)**:
  - We do not want to let [*him*]$_1$ go. [*Messi*]$_1$ is not for sale.
- **Exophora**:
  - Smoking is forbidden [*here*]$_1$.
  - [*That chair*]$_1$ is broken.

Identity noun phrase coreference

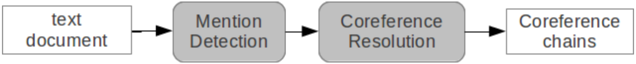
- Determining which mentions in a discourse refer to the **same real-world entity** (a coreference chain represents an entity).
- A **mention** is a noun phrase which refers to an **entity**.
- Most commonly investigated kind of coreference relation.

WE WILL FOCUS ON THEM!! For simplicity, from now, we will refer to *identity noun phrase coreference resolution* simply as *coreference resolution*.

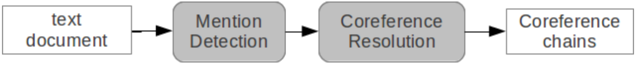
Identity noun phrase coreference

- **Included examples**:
  - [*Messi*]$_1$ is not for sale. We do not want to let [*him*]$_1$ go.
  - [*The car*]$_1$ hit a tree. [*The vehicle*]$_1$ was found one day later.
  - We do not want to let [*him*]$_1$ go. [*Messi*]$_1$ is not for sale.
- **Excluded examples**:
  - [*Laporta warned Chelsea off Messi*]$_1$. [*This warning*]$_1$ generated discouragement in Chelsea. (non-entity endophora)
  - Smoking is forbidden [*here*]$_1$. (exophora)
  - [*Every dog*]$_1$ has [*its*]$_1$ day. (bound variable)
  - The boy entered [*the room*]$_1$. The [*door*]$_1$ closed automatically. (non-identity coreference)

General methodology of a coreference solver



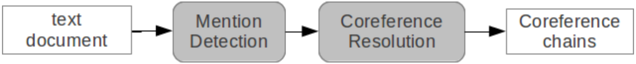
General methodology of a coreference solver



Preprocessing:

- **Mention detection**:
  - Detects the boundaries of the mentions in the input text.
  - *system mentions* vs *true mentions*
  - $\mathbf{m} = (m_1, m_2, \ldots, m_n)$ ordered as found in the document.

General methodology of a coreference solver



Coreference resolution:

- find the coreference chains.
- **Heuristic-driven approaches**: based on the centering theory of the discourse [Grosz et al., 83, 95]. See details in [Walker et al., 98].
- **ML-based approaches**: WE WILL FOCUS ON THEM!

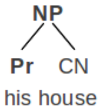
Mention detection

Mention detection

- Preprocess: POS-tagging, NERC and parsing.
- Recursively visiting the parse tree, accept the following as mention
  - Pronouns (filter out pleonastic pronouns, e.g., **It** *is raining*)
  - Proper names
  - Maximal noun phrase (NP) projections, with some exceptions.
  - Coordinated NPs

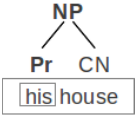
Mention detection

Examples of maximal NP projections:



Mention detection

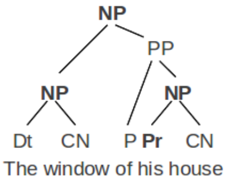
Examples of maximal NP projections:



keep the pronoun

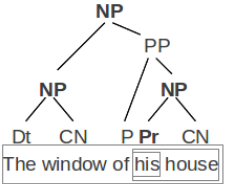
Mention detection

Examples of maximal NP projections:



Mention detection

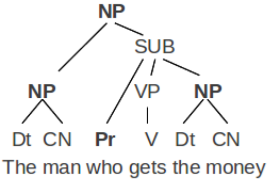
Examples of maximal NP projections:



drop out NPs sharing the same head

Mention detection

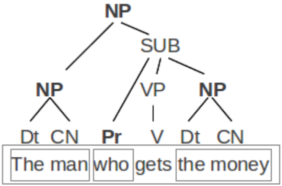
Examples of maximal NP projections:



essential subordinate clause

Mention detection

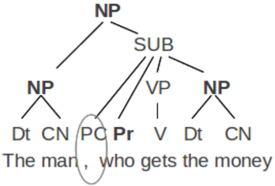
Examples of maximal NP projections:



essential subordinate clause keep NPs sharing the same head

Mention detection

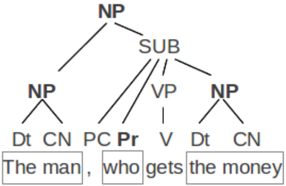
Examples of maximal NP projections:



non-essential subordinate clause

Mention detection

Examples of maximal NP projections:



non-essential subordinate clause: do not keep the maximal NP

ML-based coreference resolution

Mention-Pair model

Entity-Mention model

Ranking models

Mention-Pair model

- Examples: (m_i, m_j) classified as CO/NC.

Mention-Pair model

- Examples: (m_i, m_j) classified as CO/NC.
- **Two steps:**
  - **Learn a classifier of mention pairs**. Ex:
    - Decision Trees [McCarthy & Lehnert, 95], [Soon et al., 01]
    - Rule induction (RIPPER) [Ng & Cardie, 02]
    - Maximum Entropy [Denis & Baldrige, 07], [Ji et al., 05]
    - SVMs [Yang et al., 06]
  - **Generate chains**. Ex:
    - Closest-first strategy [Soon et al., 01]
    - Best-first strategy [Ng & Cardie, 02][Bengtson & Roth, 08]
    - Clustering [Klenner & Ailloud 2008]...
    - Global optimization (ILP) [Klenner, 07][Finkel & Manning, 08]
    - Graph partitioning [McCallum & Wellner,05][Nicolae & Nicolae, 06][Sapena et al, 10]...

Learn a mention pair classifier

Strategies for creating training examples from anotated chains

- **Closest antecendent** Given a mention m_j,
  - A positive example, $(m_i, m_j) = CO$, generated with m_j and its closest preceding antecedent m_i.
  - A set of negative examples, $\{(m_k, m_j) = NC\}$, generated with m_j and any other preceding non-antecedent m_k occurring between m_i and m_j.

Learn a mention pair classifier

Strategies for creating training examples from anotated chains

- **Closest antecendent** Given a mention m_j,
  - A positive example, $(m_i, m_j) = CO$, generated with m_j and its closest preceding antecedent m_i.
  - A set of negative examples, $\{(m_k, m_j) = NC\}$, generated with m_j and any other preceding non-antecedent m_k occurring between m_i and m_j.

Ex:

[This] is the third time [[president Obama] and [[his]wife]] come to [Spain]. [Obama] will meet with [Spanish president].

Learn a mention pair classifier

Strategies for creating training examples from anotated chains

- **Closest antecedent** Given a mention m_j,
  - A positive example, $(m_i, m_j) = CO$, generated with m_j and its closest preceding antecedent m_i.
  - A set of negative examples, $\{(m_k, m_j) = NC\}$, generated with m_j and any other preceding non-antecedent m_k occurring between m_i and m_j.

Ex:

[This] is the third time [[president Obama] and [[his]wife]] come to [Spain]. [Obama] will meet with [Spanish president].

CO: (president Obama, his), (his, Obama)

NC: (Spain, Obama) (his wife, Obama)

Learn a mention pair classifier

Strategies for creating training examples from anotated chains

- **Closest antecedent** Given a mention m_j,
  - A positive example, $(m_i, m_j) = CO$, generated with m_j and its closest preceding antecedent m_i.
  - A set of negative examples, $\{(m_k, m_j) = NC\}$, generated with m_j and any other preceding non-antecedent m_k occurring between m_i and m_j.

  Ex:

  [This] is the third time [[president Obama] and [[his]wife]] come to [Spain]. [Obama] will meet with [Spanish president].

  CO: (president Obama, his), (his, Obama)

  NC: (Spain, Obama) (his wife, Obama)

- classifier biased to select the closest antecedent
- **problem:** m_i is a pronoun

Learn a mention pair classifier

Strategies for creating training examples from anotated chains

- **Most confident closest antecedent** Given a mention m_j,
  - A positive example, $(m_i, m_j) = CO$, generated with m_j and ...
    - its closest non-pronominal antecedent m_i, if m_j is non-pronominal
    - its closest preceding antecedent m_i, otherwise
  - A set of negative examples, $\{(m_k, m_j) = NC\}$, is generated as in **closest antecedent strategy**.

Learn a mention pair classifier

Strategies for creating training examples from anotated chains

- **Most confident closest antecedent** Given a mention m_j,
  - A positive example, $(m_i, m_j) = CO$, generated with m_j and ...
    - its closest non-pronominal antecedent m_i, if m_j is non-pronominal
    - its closest preceding antecedent m_i, otherwise
  - A set of negative examples, $\{(m_k, m_j) = NC\}$, is generated as in **closest antecedent strategy**.
- Ex:

  [This] is the third time [[president Obama] and [[his]wife]] come to [Spain]. [Obama] will meet with [Spanish president].

Learn a mention pair classifier

Strategies for creating training examples from anotated chains

- **Most confident closest antecedent** Given a mention m_j,
  - A positive example, $(m_i, m_j) = CO$, generated with m_j and ...
    - its closest non-pronominal antecedent m_i, if m_j is non-pronominal
    - its closest preceding antecedent m_i, otherwise
  - A set of negative examples, $\{(m_k, m_j) = NC\}$, is generated as in **closest antecedent strategy**.
- Ex:

  [This] is the third time [[president Obama] and [[his]wife]] come to [Spain]. [Obama] will meet with [Spanish president].

  CO: (president Obama, his), (president Obama, Obama)

  NC: (Spain, Obama), (his wife, Obama), (president Obama and his wife, Obama)

Learn a mention pair classifier

Strategies for creating training examples from anotated chains

- **Most confident closest antecedent** Given a mention m_j,
  - A positive example, $(m_i, m_j) = CO$, generated with m_j and ...
    - its closest non-pronominal antecedent m_i, if m_j is non-pronominal
    - its closest preceding antecedent m_i, otherwise
  - A set of negative examples, $\{(m_k, m_j) = NC\}$, is generated as in **closest antecedent strategy**.
- Ex:

  [This] is the third time [[president Obama] and [[his]wife]] come to [Spain]. [Obama] will meet with [Spanish president].

  CO: (president Obama, his), (president Obama, Obama)

  NC: (Spain, Obama), (his wife, Obama), (president Obama and his wife, Obama)
- classifier biased to select the most confident closest antecedent

Learn a mention pair classifier

Strategies for creating training examples from anotated chains

- **All antecedents**
  - Positive examples, $\{(m_i, m_j) = CO\}$, generated with any pair m_i and m_j anotated in the same chain
  - Negative examples, generated from the rest of mention pairs.

Learn a mention pair classifier

Strategies for creating training examples from anotated chains

- **All antecedents**
  - Positive examples, $\{(m_i, m_j) = CO\}$, generated with any pair m_i and m_j anotated in the same chain
  - Negative examples, generated from the rest of mention pairs.
- Ex:

  [This] is the third time [[president Obama] and [[his]wife]] come to [Spain]. [Obama] will meet with [Spanish president].

Learn a mention pair classifier

Strategies for creating training examples from anotated chains

- **All antecedents**
  - Positive examples, $\{(m_i, m_j) = CO\}$, generated with any pair m_i and m_j anotated in the same chain
  - Negative examples, generated from the rest of mention pairs.
- Ex:

  [This] is the third time [[president Obama] and [[his]wife]] come to [Spain]. [Obama] will meet with [Spanish president].

  CO: (president Obama, Obama), (his, Obama), (president Obama, his)

  NC: (This, president Obama), (This, his), (This, his wife), (This, Spain), ..., ((Spain, Spanish president), (Obama, Spanish president)

Learn a mention pair classifier

Filtering training examples

- Reduce the number of training examples trying to increase precision
- Apply some restrictions to set $\{(m_i, m_j)\}$
  - if (m_i, m_j) involves a pronoun and violates gender and number agreement, remove the example.
  - if m_j is a pronoun, restrict the search of m_i in 3 sentences before.
  - ...
- Apply example selection strategies to set $NC = \{(m_i, m_j)\}$
  - select $e^- = (m_i, m_j) \in NC$ if $\exists e^+ = (m_s, m_t) \in CO$: $d(e^+, e^-) < threshold$
  - ...

Learn a mention pair classifier

Mentions characterization (examples of feature functions)

Type	Feature	Description
Structural	DIST_SEN_k	distance in sentences is k: y,n
	DIST_SEN_>2	distance in sentences greater than 2: y,n
	DIST_MEN_k	distance in mentions is k: y,n
	DIST_MEN_>2	distance in mentions greater than 2: y,n
	APPOSITIVE	One mention in apposition with the other: y,n
Lexical	STR_MATCH	String matching: y,n
	ALIAS	One mention is an alias of the other: y,n,u
Morphological	NUMBER	The number of both mentions match: y,n,u
	GENDER	The gender of both mentions match: y,n,u
Syntactic	DEF_NP	m_j is a definitive NP: y,n
	DEM_NP	m_j is a demonstrative NP: y,n
Semantical	SEMCLASS	Semantic class match: y,n,u
	ANIMACY	Animacy of both mentions match: y,n

Learn a mention pair classifier

Mentions characterization (examples of feature functions)

Type	Feature	Description
Structural	DIST_SEN_k	distance in sentences is k: y,n
	DIST_SEN_>2	distance in sentences greater than 2: y,n
	DIST_MEN_k	distance in mentions is k: y,n
	DIST_MEN_>2	distance in mentions greater than 2: y,n
	APPOSITIVE	One mention in apposition with the other: y,n
Lexical	STR_MATCH	String matching: y,n
	ALIAS	One mention is an alias of the other: y,n,u
Morphological	NUMBER	The number of both mentions match: y,n,u
	GENDER	The gender of both mentions match: y,n,u
Syntactic	DEF_NP	m_j is a definitive NP: y,n
	DEM_NP	m_j is a demonstrative NP: y,n
Semantical	SEMCLASS	Semantic class match: y,n,u
	ANIMACY	Animacy of both mentions match: y,n

Learn a mention pair classifier

Mentions characterization (examples of feature functions)

Type	Feature	Description
Structural	DIST_SEN_k	distance in sentences is k: y,n
	DIST_SEN_>2	distance in sentences greater than 2: y,n
	DIST_MEN_k	distance in mentions is k: y,n
	DIST_MEN_>2	distance in mentions greater than 2: y,n
	APPOSITIVE	One mention in apposition with the other: y,n
Lexical	STR_MATCH	String matching: y,n
	ALIAS	One mention is an alias of the other: y,n,u
Morphological	NUMBER	The number of both mentions match: y,n,u
	GENDER	The gender of both mentions match: y,n,u
Syntactic	DEF_NP	m_j is a definitive NP: y,n
	DEM_NP	m_j is a demonstrative NP: y,n
Semantical	SEMCLASS	Semantic class match: y,n,u
	ANIMACY	Animacy of both mentions match: y,n

Learn a mention pair classifier

Mentions characterization (examples of feature functions)

Type	Feature	Description
Structural	DIST_SEN_k	distance in sentences is k: y,n
	DIST_SEN_>2	distance in sentences greater than 2: y,n
	DIST_MEN_k	distance in mentions is k: y,n
	DIST_MEN_>2	distance in mentions greater than 2: y,n
	APPOSITIVE	One mention in apposition with the other: y,n
Lexical	STR_MATCH	String matching: y,n
	ALIAS	One mention is an alias of the other: y,n,u
Morphological	NUMBER	The number of both mentions match: y,n,u
	GENDER	The gender of both mentions match: y,n,u
Syntactic	DEF_NP	m_j is a definitive NP: y,n
	DEM_NP	m_j is a demonstrative NP: y,n
Semantical	SEMCLASS	Semantic class match: y,n,u
	ANIMACY	Animacy of both mentions match: y,n

Learn a mention pair classifier

Mentions characterization (examples of feature functions)

Type	Feature	Description
Structural	DIST_SEN_k	distance in sentences is k: y,n
	DIST_SEN_>2	distance in sentences greater than 2: y,n
	DIST_MEN_k	distance in mentions is k: y,n
	DIST_MEN_>2	distance in mentions greater than 2: y,n
	APPOSITIVE	One mention in apposition with the other: y,n
Lexical	STR_MATCH	String matching: y,n
	ALIAS	One mention is an alias of the other: y,n,u
Morphological	NUMBER	The number of both mentions match: y,n,u
	GENDER	The gender of both mentions match: y,n,u
Syntactic	DEF_NP	m_j is a definitive NP: y,n
	DEM_NP	m_j is a demonstrative NP: y,n
Semantical	SEMCLASS	Semantic class match: y,n,u
	ANIMACY	Animacy of both mentions match: y,n

Learn a mention pair classifier

Mentions characterization (examples of feature functions)

Type	Feature	Description
Structural	DIST_SEN_k	distance in sentences is k: y,n
	DIST_SEN_>2	distance in sentences greater than 2: y,n
	DIST_MEN_k	distance in mentions is k: y,n
	DIST_MEN_>2	distance in mentions greater than 2: y,n
	APPOSITIVE	One mention in apposition with the other: y,n
Lexical	STR_MATCH	String matching: y,n
	ALIAS	One mention is an alias of the other: y,n,u
Morphological	NUMBER	The number of both mentions match: y,n,u
	GENDER	The gender of both mentions match: y,n,u
Syntactic	DEF_NP	m_j is a definitive NP: y,n
	DEM_NP	m_j is a demonstrative NP: y,n
Semantical	SEMCLASS	Semantic class match: y,n,u
	ANIMACY	Animacy of both mentions match: y,n

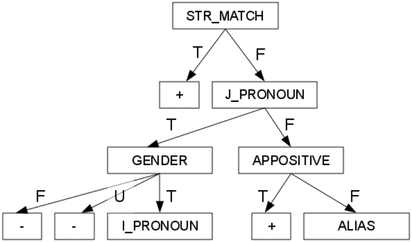
Learn a mention pair classifier

Dataset for training

Pair	DIST_SEN_0	...	DIST_MEN_>2	APPOSITIVE	STR_MATCH	ALIAS	NUMBER	GENDER	DEF_NP	DEM_NP	SEMCLASS	ANIMACY	Corefer?
m_1, m_2	y		n	n	n	u	n	y	n	n	y	y	N
m_1, m_3	y		n	n	n	n	u	n	n	n	n	n	N
m_1, m_4	n		n	n	y	y	y	u	n	n	y	y	Y
m_1, m_3	n		y	n	n	n	y	n	n	n	y	y	N
⋮													
m_i, m_j	n		y	n	n	n	y	y	n	n	y	y	Y

Learn a mention pair classifier

Decision Tree [McCarthy & Lehnert, 95], [Soon et al., 01]
Ex:



Learn a mention pair classifier

Maximum Entropy [Denis & Baldrige, 07], [Ji et al., 05]
Ex:

$$P(CO|x_{ij}) = \frac{\exp \sum_k \lambda_k f_k(x_{ij}, CO)}{\sum_c \exp \sum_k \lambda_k f_k(x_{ij}, c)}$$

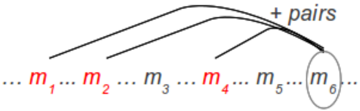
Ex:

$$f_s(x_{ij}, CO) = \begin{cases} 1 & APPOSITIVE(x_{ij}) \quad and \quad corefer(x_{ij}) \\ 0 & otherwise \end{cases}$$

Maximum likelihood estimation of parameters λ_i (e.g. Improved Iterative scaling [Della Pietra et al., 96])

Generate chains

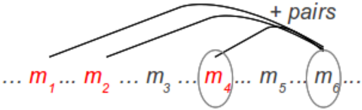
Closest-first strategy [Soon et al., 01]



if a probabilistic classifier is used then define a threshold above which a pair is considered coreferent (i.e, +pairs)

Generate chains

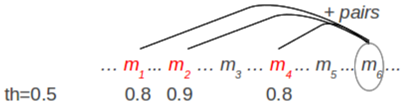
Closest-first strategy [Soon et al., 01]



for a given m_j, select as antecedent the closest preceding m_k from the +pairs

Generate chains

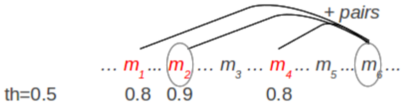
Best-first strategy [Ng & Cardie, 02][Bengtson & Roth, 08]



aims to improve the Precision of closest-first clustering by taking profit of the probabilities of the +pairs

Generate chains

Best-first strategy [Ng & Cardie, 02][Bengtson & Roth, 08]

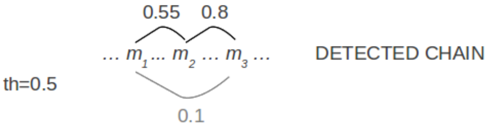


for a given m_j, select as antecedent the most probable precedent m_k from the $+$pairs

Generate chains

Drawback due to the use of closest-first or best-first strategies:

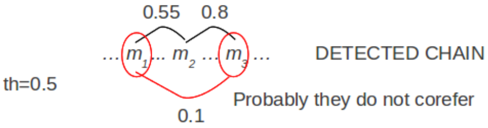
- They only take profit of individual +pairs decisions of the mention-pair classifier



Generate chains

Drawback due to the use of closest-first or best-first strategies:

- They only take profit of individual +pairs decisions of the mention-pair classifier



Generate chains

possible solution: take profit of groups of decisions of the mention-pair classifier.

Generate chains

possible solution: take profit of groups of decisions of the mention-pair classifier.

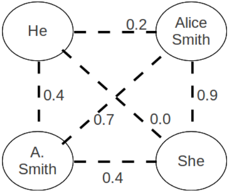
Approaches based on:

- Clustering [Klenner & Ailloud 2008]...
- Global optimization (ILP) [Klenner, 07][Finkel & Manning, 08]
- Graph partitioning algorithms [McCallum & Wellner,05][Nicolae & Nicolae, 06][Sapena et al, 10]...

Generate chains

Graph-partitioning algorithms

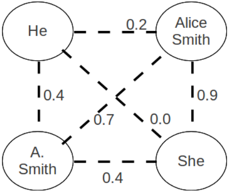
- Ex: Alice Smith ... A. Smith ... He ... She ...



Generate chains

Graph-partitioning algorithms

- Ex: Alice Smith ... A. Smith ... He ... She ...

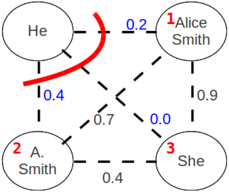


- Find the most appropriated partition P in order to isolate the groups that represent independent entities.

Generate chains

Graph-partitioning algorithms

Ex: Alice Smith ... A. Smith ... He ... She ...



- Find the most appropriated partition P in order to isolate the groups that represent independent entities.
- P can be learned from training data

Generate chains

Graph-partitioning algorithms

Ex: Constraint relaxation [Sapena et al., 10]

A more flexible approach

- each mentions m_i is a vertex in the graph
- each pair of mentions (m_i, m_j) is connected by an edge e_{ij}
- each edge e_{ij} is weighted by w_{ij}

$$w_{ij} = \sum_{k \in C_{ij}} \lambda_k$$

C_{ij}: set of constraints that restrict the compatibility between m_i and m_j

λ_k: weight associated to the constraint k

λ_k and w_{ij} can be negative

Generate chains

Graph-partitioning algorithms

Ex: Constraint relaxation [Sapena et al., 10]

A more flexible approach

- each mentions m_i is a vertex in the graph
- each pair of mentions (m_i, m_j) is connected by an edge e_{ij}
- each edge e_{ij} is weighted by w_{ij}

$$w_{ij} = \sum_{k \in C_{ij}} \lambda_k$$

C_{ij}: set of constraints that restrict the compatibility between m_i and m_j

λ_k: weight associated to the constraint k

λ_k and w_{ij} can be negative

Generate chains

Graph-partitioning algorithms

Ex: Constraint relaxation [Sapena et al., 10]

- C_{ij}: A Decision Tree (DT) is learned from mention pairs. Each rule in the DT is a constraint. C_{ij} is the set of constraints satisfied between m_i and m_j.

Generate chains

Graph-partitioning algorithms

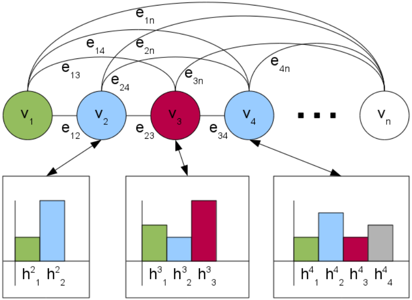
Ex: Constraint relaxation [Sapena et al., 10]

- C_{ij}: A Decision Tree (DT) is learned from mention pairs. Each rule in the DT is a constraint. C_{ij} is the set of constraints satisfied between m_i and m_j.
- $\lambda_k = P_k - \alpha$

Generate chains

Graph-partitioning algorithms

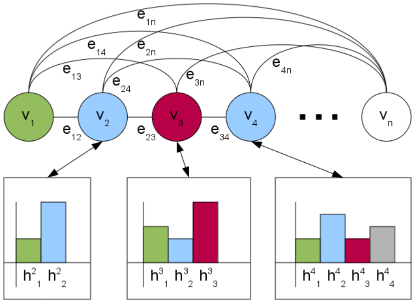
Ex: Constraint relaxation [Sapena et al., 10]



Generate chains

Graph-partitioning algorithms

Ex: Constraint relaxation [Sapena et al., 10]



$$h_l^i = argmax_{h_l^i} \sum_{l \in |L_i|} h_l^i S_{il}$$
$$S_{il} = \sum_{e_{ij}} w_{ij} x h_l^j$$

Generate chains

Graph-partitioning algorithms

Ex: Constraint relaxation [Sapena et al., 10]

$H = H(0)$;
repeat
  for *each* m_i **do**
    for *each* $l \in |L_i|$ **do**
      $S_{il} = \sum_{e_{ij}} w_{ij} x h_l^j$
    end
    Normalize all S_{il} to $[-1, 1]$;
    for *each* $l \in |L_i|$ **do**
      $h_l^i(t+1) = (h_l^i(t) x (1 + S_{il})) / (\sum_{k \in |L_i|} h_l^k(t) x (1 + S_{ik}))$
    end
  end
until $H(t+1) \approx H(t)$;

Generate chains

Graph-partitioning algorithms

Ex: Constraint relaxation [Sapena et al., 10]

$H = H(0)$;

repeat

    for *each* m_i **do**

        for *each* $l \in |L_i|$ **do**

            $S_{il} = \sum_{e_{ij}} w_{ij} x h_l^j$

        end

        Normalize all S_{il} to $[-1, 1]$;

        for *each* $l \in |L_i|$ **do**

            $h_l^i(t+1) = (h_l^i(t) x (1 + S_{il})) / (\sum_{k \in |L_i|} h_l^k(t) x (1 + S_{ik}))$

        end

    end

until $H(t+1) \approx H(t)$;

Generate chains

Graph-partitioning algorithms

Ex: Constraint relaxation [Sapena et al., 10]

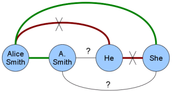
- $H(0)$: the probability of m_i, being in a new partition is:
  - m_i is not a pronoun: the probability of m_i, being in a new partition is double than the probabilities for the rest of coreferent candidates. (realistic situation: majority of mentions are singletons)

$$h_l^i = \frac{1}{|L_i|+2} \quad \forall_{l \in [0,|L_i|-1]} \qquad h_{|L_i|}^i = \frac{2}{|L_i|+2}$$

  - m_i is a pronoun: equiprobable.

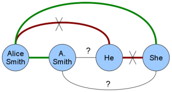
Drawbacks of the Mention-Pair model

Lack of information.

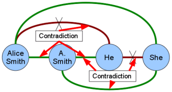


Drawbacks of the Mention-Pair model

Lack of information.

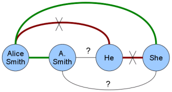


Contradictions in classification.

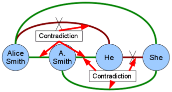


Drawbacks of the Mention-Pair model

Lack of information.



Contradictions in classification.

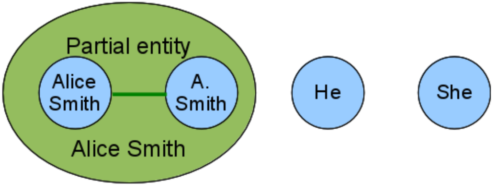


- Generating chains by graph partitioning, ILP or clustering methods adress this problem within the mention-pair model paradigm.
- Entity-mention model and ranking models are different perspectives to deal with the problem.

Entity-Mention model

Entities characterization

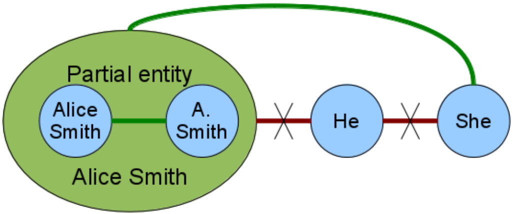
- Feature functions used for the Mention-Pair models
- New feature functions implying mention groups



Entity-Mention model

Entities characterization

- Feature functions used for the Mention-Pair models
- New feature functions implying mention groups



Entity-Mention model

Examples:

- Constraint relaxation [Sapena, 12]
- Global optimization [Luo et al., 04]
- Clustering [Ng., 08]

Entity-Mention model

Examples:

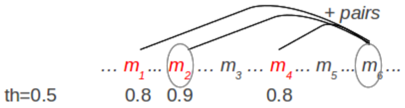
- Constraint relaxation [Sapena, 12]
- Global optimization [Luo et al., 04]
- Clustering [Ng., 08]

Pros: improved expressiveness

Cons: the results achieved are not particularly encouraging

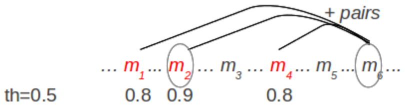
Ranking models

Mention-pair models take profit of independent mention pair decisions between m_i and each possible antecedent.



Ranking models

Mention-pair models take profit of independent mention pair decisions between m_i and each possible antecedent.



Ranking models take profit of decisions between m_i and all its possible antecedents.

Ranking models

Ex: rankers [Denis and Baldrige, 07]

- Learn a ranker from examples
- Example $= (m_i, \alpha_i, A_i)$, where α_i is the first antecedent of m_i and A_i is the set of non-antecedents in a window of 4 sentences around α_i (including its own sentence).
- Exponential model:

$$P(\alpha_i|m_i) = \frac{\exp \sum_k \lambda_k f_k(m_i, \alpha_i)}{\sum_{m_s \in A_i \cup \{m_i\}} \exp \sum_k \lambda_k f_k(m_i, m_s)}$$

Ranking models

Ex: rankers [Denis and Baldrige, 07]

- Learn a ranker from examples
- Example $= (m_i, \alpha_i, A_i)$, where α_i is the first antecedent of m_i and A_i is the set of non-antecedents in a window of 4 sentences around α_i (including its own sentence).
- Exponential model:

$$P(\alpha_i | m_i) = \frac{\exp \sum_k \lambda_k f_k(m_i, \alpha_i)}{\sum_{m_s \in A_i \cup \{m_i\}} \exp \sum_k \lambda_k f_k(m_i, m_s)}$$

- Resolution: A_i is the set of the preceding mentions of m_i in a window of 4 sentences

Ranking models

Examples:

for mentions [Yang et al., 03][Denis and Baldrige, 08]

for partial entities [Rahman and Ng, 09]

Ranking models

Examples:

- **for mentions** [Yang et al., 03][Denis and Baldrige, 08]
- **for partial entities** [Rahman and Ng, 09]

Pros: take profit of decisions involving all the candidate antecedents.

Cons: always pick an antecedent from the candidates, although the mention in course is not anaphoric.

Ranking models

Examples:

> **for mentions** [Yang et al., 03][Denis and Baldrige, 08]
>
> **for partial entities** [Rahman and Ng, 09]

Pros: take profit of decisions involving all the candidate antecedents.
Cons: always pick an antecedent from the candidates, although the mention in course is not anaphoric.
(a classifier of anaphoricity improves the results)

Comparison

CoNLL 2011 shared task [Pradhan et al., 11]

System	MD	MUC	B-CUBED	CEAF$_m$	CEAF$_e$	BLANC	Official	
	F	F^1	F^2	F	F^3	F	$\frac{F^1+F^2+F^3}{3}$	
lee	70.70	59.57	68.31	56.37	45.48	73.02	57.79	Handcrafted rules
sapena	43.20	59.55	67.09	53.51	41.32	71.10	55.99	
chang	64.28	57.15	68.79	54.40	41.94	73.71	55.96	
nugues	68.96	58.61	65.46	51.45	39.52	71.11	54.53	
santos	65.45	56.65	65.66	49.54	37.91	69.46	53.41	
song	67.26	59.95	63.23	46.29	35.96	61.47	53.05	
stoyanov	67.78	58.43	61.44	46.08	35.28	60.28	51.92	
sobha	64.23	50.48	64.00	49.48	41.23	63.28	51.90	
kobdani	61.03	53.49	65.25	42.70	33.79	62.61	51.04	
zhou	62.31	48.96	64.07	47.53	39.74	64.72	50.92	
charton	64.30	52.45	62.10	46.22	36.54	64.20	50.36	
yang	63.93	52.31	62.32	46.55	35.33	64.63	49.99	
hao	64.30	54.47	61.01	45.07	32.67	65.35	49.38	
xinxin	61.92	46.62	61.93	44.75	36.23	64.27	48.46	
zhang	61.13	47.28	61.14	44.46	35.19	65.21	48.07	
kummerfeld	62.72	42.70	60.29	45.35	38.32	59.91	47.10	
zhekova	48.29	24.08	61.46	40.43	35.75	53.77	40.43	
irwin	26.67	19.98	50.46	31.68	25.21	51.12	31.28	

Comparison

CoNLL 2011 shared task [Pradhan et al., 11]

System	MD	MUC	B-CUBED	CEAF$_m$	CEAF$_e$	BLANC	Official	
	F	F^1	F^2	F	F^3	F	$\frac{F^1+F^2+F^3}{3}$	
lee	70.70	59.57	68.31	56.37	45.48	73.02	57.79	
sapena	43.20	59.55	67.09	53.51	41.32	71.10	55.99	DT+constrain relaxation
chang	64.28	57.15	68.79	54.40	41.94	73.71	55.96	
nugues	68.96	58.61	65.46	51.45	39.52	71.11	54.53	
santos	65.45	56.65	65.66	49.54	37.91	69.46	53.41	
song	67.26	59.95	63.23	46.29	35.96	61.47	53.05	
stoyanov	67.78	58.43	61.44	46.08	35.28	60.28	51.92	
sobha	64.23	50.48	64.00	49.48	41.23	63.28	51.90	
kobdani	61.03	53.49	65.25	42.70	33.79	62.61	51.04	
zhou	62.31	48.96	64.07	47.53	39.74	64.72	50.92	
charton	64.30	52.45	62.10	46.22	36.54	64.20	50.36	
yang	63.93	52.31	62.32	46.55	35.33	64.63	49.99	
hao	64.30	54.47	61.01	45.07	32.67	65.35	49.38	
xinxin	61.92	46.62	61.93	44.75	36.23	64.27	48.46	
zhang	61.13	47.28	61.14	44.46	35.19	65.21	48.07	
kummerfeld	62.72	42.70	60.29	45.35	38.32	59.91	47.10	
zhekova	48.29	24.08	61.46	40.43	35.75	53.77	40.43	
irwin	26.67	19.98	50.46	31.68	25.21	51.12	31.28	

closest-first

best-first

clustering

ILP

rankers

References

- B.J. Grosz, A.K. Joshi and S. Weinstein, **Providing a unified account of definite noun phrases in discourse**. Proceedings of ACL, 1983.
- B.J. Grosz, A.K. Joshi and S. Weinstein, **Centering: A framework for modeling the local coherence of discourse**. Computational Linguistics, 21(2), 1995.
- M. Walker, A. Joshi, and E. Prince, editors. 1998. **Centering theory in discourse**. Oxford University Press
- W.M. Soon, H.T. Ng and D.C.Y. Lim, **A machine learning approach to coreference resolution of noun phrases**. Computational Linguistics, 27(4), 2001
- V. Ng and C. Cardie, **Improving machine learning approaches to coreference resolution**. Proceedings of ACL, 2002

References

- P. Denis and J. Baldridge, **A ranking approach to pronoun resolution**. Proceedings of IJCAI, 2007.
- P. Denis and J. Baldridge, **Specialized models and ranking for coreference resolution**. Proceedings of EMNLP, 2008.
- V. Ng, **Supervised noun phrase coreference research: The First Fifteen Years**. Proceedings of ACL, 2010.
- E. Sapena, L. Padró and J. Turmo. 2010. **A global relaxation labeling approach to coreference resolution**. Proceedings of COLING 2010
- E. Sapena. 2012. **A constraint-based hypergraph partitioning approach to coreference resolution**. PhD. Thesis
- S. Pradhan, L. Ramshaw, M. Marcus, M. Palmer, R. Weischedel and N. Xue. 2011 **CoNLL-2011 Shared Task: Modeling Unrestricted Coreference in OntoNotes**