

Reference Resolution

Regina Barzilay

February 23, 2004

Announcements

3/3 — first part of the projects

Example topics

- Segmentation
- Identification of discourse structure
- Summarization
- Anaphora resolution
- Cue phrase selection

Reference Resolution

Captain Farragut was a good seaman, worthy of the frigate he commanded. His vessel and he were one. He was the soul of it.

- Coreference resolution: {the frigate, his vessel, it}
- Anaphora resolution: {his vessel, it}

Coreference is a harder task!

Last Time

- Symbolic Multi-Strategy Anaphora Resolution
(Lappin&Leass, 1994)
- Clustering-based Coreference Resolution
(Cardie&Wagstaff, 1999)
- Supervised ML Coreference Resolution + Clustering
(Soon et al, 2001), (Ng&Cardie, 2002)

Features (Soon et al, 2001)

- distance in sentences between anaphora and antecedent?
- antecedent in a pronoun?
- weak string identity between anaphora and antecedent?
- anaphora is a definite noun phrase?
- anaphora is a demonstrative pronoun?
- number agreement between anaphora and antecedent
- semantic class agreement anaphora and antecedent
- gender agreement between anaphora and antecedent
- anaphora and antecedent are both proper names?
- an alias feature
- an appositive feature

Observations

(Ng&Cardie'2002)

0,76,83,C,D,C,D,D,D,D,D,I,I,C,I,I,D,N,N,D,C,D,D,N,N,N,N,N,C,Y,
Y,D,D,D,C,0,D,D,D,D,D,D,D,1,D,D,C,N,Y,D,D,D,20,20,D,D,-.
0,75,83,C,D,C,D,D,D,C,D,I,I,C,I,I,C,N,N,D,C,D,D,N,N,N,N,N,C,Y,
Y,D,D,D,C,0,D,D,D,D,D,D,C,1,D,D,C,Y,Y,D,D,D,20,20,D,D,+.
0,74,83,C,D,C,D,D,D,D,D,I,I,C,I,I,D,N,N,D,C,D,D,N,N,N,N,N,C,Y,
Y,D,D,D,C,0,D,D,D,D,D,D,D,1,D,D,C,N,Y,D,D,D,20,20,D,D,-.

Classification Rules

+ 786 59 IF SOON-WORDS-STR = C
+ 73 10 IF WNCLASS = C PROPER-NOUN = D NUMBERS = C SENTNUM <= 1 PRO-RESOLVE = C ANIMACY = C
+ 40 8 IF WNCLASS = C CONSTRAINTS = D PARANUM <= 0 PRO-RESOLVE = C
+ 16 0 IF WNCLASS = C CONSTRAINTS = D SENTNUM <= 1 BOTH-IN-QUOTES = I APPOSITIVE = C
+ 17 0 IF WNCLASS = C PROPER-NOUN = D NUMBERS = C PARANUM <= 1 BPRONOUN-1 = Y AGREEMENT = C CONSTRAINTS = C BOTH-PRONOUNS = C
+ 38 24 IF WNCLASS = C PROPER-NOUN = D NUMBERS = C SENTNUM <= 2 BOTH-PRONOUNS = D AGREEMENT = C SUBJECT-2 = Y
+ 36 8 IF WNCLASS = C PROPER-NOUN = D NUMBERS = C BOTH-PROPER-NOUNS = C
+ 11 0 IF WNCLASS = C CONSTRAINTS = D SENTNUM <= 3 SUBJECT-1 = Y SUBJECT-2 = Y SUBCLASS = D IN-QUOTE-2 = N BOTH-DEFINITES = I

Observations

- Feature selection plays an important role in classification accuracy: MUC-6 62.6% (Soon et al., 2001) → Ng&Cardie, 2002) 69.1%
- Clustering operates over the results of hard clustering, which may negatively influence the final results
- Machine learning techniques rely on large amounts of annotated data: 30 texts
- All the methods are developed on the same corpus of newspaper articles

Today

- Minimizing amounts of training data:
 - Co-training
 - Weakly-supervised learning
- Hobbs' algorithm
- Anaphora resolution in dialogs

Co-training

(Blum&Mitchell, 1998)

1. Given a small amount of training data, train two classifiers based on orthogonal set of features
2. Add to training set n instances on which both classifiers agree
3. Retrain both classifiers on the extended set
4. Return to step 2

Co-training for Coreference

Coreference does not support natural split of features

Algorithm for feature splitting

- Train a classifier on each feature separately
- Select the best feature and assign it to the first view, and the second best feature assign to the second view
- Iterate over the remaining feature, and add them to one of the views

Separate training for each reference type (personal pronouns, possessives, . . .)

Results

Improvements for some types of references

- Definite noun phrases: from 19% to 28% (2000 training instances)
- No improvements for possessives, proper names and possessive pronouns

Study of learning curves

- Personal and possessive pronoun can be trained from very small training data (100 instances)
- Other types of references require large amounts of training data

Anaphora In Spoken Dialogue

Differences between spoken and written text

- High frequency of anaphora
- Presence of “Vague anaphora”
(Eckert&Strube’2000) 33%
- Presence of non-NP-antecedents
(Byron&Allen’1998) TRAINS93: 50%
(Eckert&Strube’2000) SwitchBoard: 22%
- Presence of repairs, disfluences, abandoned utterances and so on...

Example of Dialog

A1: ..[he]_i's nine months old ...

A2: ..[He]_i likes to dig around a little bit.

A3: ..[His mother]_i mother comes in and says, why did you let [him]_i [plays in the dirt]_j.

A4: I guess [[he]_i's enjoying himself]_k.

B5: [That]_k's right.

B6: [It]_j's healthy ...

Abstract Referents

(Webber, 1988)

(A0) Each Fall, penguins migrate to Fiji.

(A1) *That's* where they wait out the winter.

(A2) *That's* when it's cold even for them.

(A3) *That's* why I'm going there next month.

(A4) *It* happens just before the eggs hatch.

Abstract Referents

- Webber (1990): each discourse unit produces a pseudo discourse entity — “proxy for its propositional content”
- Abstract Pronoun interpretation: requires presentation of fact referents
- Walker&Whittaker (1990): in problem-solving dialogs, people refer to aspects of the solution that were not explicitly mentioned
(Byron, 2002)
A1 Send engine to Elmira.
A2 That’s six hours.

Symbolic Approach

Pronominal Anaphora Resolution (Byron, 2002)

- Mentioned Entities — referents nouns phrases
- Activated Entities — entire sentences and nominals
- Discourse Entity attributes:
 - Input: The surface linguistic constituent
 - Type: ENGINE, PERSON, ...
 - Composition: hetero- or homogeneous
 - Specificity: individual or kind

Activated Entities

Generation of Multiple Proxies

- To load the boxcars/Loading them takes an hour (infinitive or gerund phrase)
- I think he that he's an alien (the entire clause)
- I think that he's an alien (sentential)
- If he's an alien (Subordinate clause)

Types of Speech Acts

Tell, Request, Wh-Questions, YN-Question, Confirm

(1) The highway is closed (Tell)

(2) Is the highway closed? (Y/N Question)

(3) That's right.

(4) Why is the highway closed? (WH-Q)

(5) *That's right.

Semantic Constraints

“Heavily-typed” system

- Verb Senses (selectional restrictions)
“Load them into the boxcar” (*them* has to be CARGO)
- Predicate NPs
“That’s a good route “ (*that* has to be a ROUTE)
- Predicate Adjectives
“It’s right” (*it* has to be a proposition)

Example

Engine 1 goes to Avon to get the oranges.

(TELL (MOVE :theme x :dest y :reason (LOAD :theme w)))

(the x (refers-to x ENG1))

(the y (refers-to y AVON))

(the w (refers-to w ORANGES))

So it'll get there at 3 p.m.

(ARRIVE :theme x :dest: y :time z)

“get there” requires MOVABLE-OBJECT

Evaluation

10 dialogues, 557 utterances, 180 test pronouns

- Salience-based resolution: 37%
- Adding Semantic constraints: 43%
- Adding Abstract referents: 67%
- “Smart” Search order: 72%
- Domain Independent Semantics: 51%

Knowledge-Lean Approach

(Strube&Muller'2003)

- Switchboard: 3275 sentences, 1771 turns, 16601 markables
- Data annotated with disfluency information
- “Problematic” utterances were discarded
- Approach: ML combines standard features with dialogue specific features

Features

Features induced for spoken dialogue: ante-exp-type [type of antecedent (NP, S, VP)]

ana-np-pref [preference for NP arguments]

mdist-3mf3p [the number of NP-markables between anaphora and potential antecedent]

ante-tfidf [the relative importance of the expression in the dialogues]

average-ic [information content: neg. log of the total frequency of the word divided by number of words]

Features

F-measure:

- Fem&Masc Pronoun: 17.4% baseline, 17.25%
- Third Person Neuter Pronoun: 14.68%, 19.26%
- Third Person Plural: 28.30%, 28.70%

Observations

- Coreference for speech processing is hard!
- New features for dialogue are required
- Prosodic features seems to be useful

Hobbs' Algorithm

- Task: Pronoun resolution
- Features: Fully Syntactic
- Accuracy: 82%

Example

U1: Lyn's mother is a gardener. U2: Craige likes her.

Anaphora Generation

(Reiter&Dale'1995)

- Application: Lexical choice for generation
- Framework:
Context Set $C = a_1, a_2, \dots, a_n$
Properties: $p_{k_1}, p_{k_2}, \dots, p_{k_m}$
- Goal: Distinguish Referent from the Rest

Algorithm

- Check Success: see if the contracted description picks up one entity from the context
- Choose Property: determine which properties of the referent would rule out the largest number of entities
- Extend Description: add the chosen properties to the description being constructed and remove relevant entities from the discourse.

Statistical Generation

- (Radev,1998): classification-based
- (Nenkova&McKeown,2003): HMM-based