Learning the Fine-Grained Information Status of Discourse Entities

Altaf Rahman and Vincent Ng
Human Language Technology Research Institute
The University of Texas at Dallas

Plan for the talk

- What is Information Status?
- Fine-Grained Information Status
  - Rule-based Approach
  - Learning-based Approach
  - Evaluation
- Conclusion

Plan for the talk

- What is Information Status?
- Fine-Grained Information Status
  - Rule-based Approach
  - Learning-based Approach
  - Evaluation
- Conclusion

What is Information Status (IS)?

- IS determination is the problem of partitioning discourse entities in a document into different classes on the **given-new** scale
- Theoretical notions of IS are not used consistently in the literature
  - original definition due to Prince (1981), but …
  - we adopted Nissim et al.'s (2004) notion due to the availability of a corpus annotated with IS according to their notion

Nissim et al.’s Notion of Information Status

- **3-way** classification scheme for IS
- A discourse entity is
  - **Old** to the hearer if it is known to the hearer and has been previously referred to in the dialogue

I was angry that he destroyed **my** tent.

Nissim et al.'s Notion of Information Status

- **3-way** classification scheme for IS
- A discourse entity is
  - **Old** to the hearer if it is known to the hearer and has been previously referred to in the dialogue
  - **New** if it has not been previously referred to

I saw **Jenny** going to the pub.

Nissim et al.'s Notion of Information Status

- **3-way** classification scheme for IS
- A discourse entity is
  - **Old** to the hearer if it is known to the hearer and has been previously referred to in the dialogue
  - **New** if it has not been previously referred to
  - **Mediated** if it is newly mentioned but its identity can be inferred from a previously-mentioned entity

He passed by Jan's house and saw that **the door** is painted red.
The Great Wall is situated in **China**.

Plan for the talk

- What is Information Status?
- Fine-Grained Information Status
  - Rule-based Approach
  - Learning-based Approach
  - Evaluation
- Conclusion

What is Fine-Grained IS?

- Nissim et al. (2004) subcategorize
  - **old** into 6 subtypes
  - **med** into 9 subtypes
- No subcategorization for **new**
- We define the fine-grained IS determination problem as one where we classify an NP as belonging to one of 16 subtypes
  - 6 **old** subtypes, 9 **med** subtypes, and **new**

The 16 subtypes

- IS subtypes
  - Old (6)
    - identity
    - event
    - general
    - generic
    - ident_generic
    - relative
  - New (1)
  - Mediated (9)
    - general
    - event
    - bound
    - part
    - situation
    - set
    - poss
    - func_value
    - aggregation

The Old Subtypes

- Old (6)
  - **identity**
  - event
  - general
  - generic
  - ident_generic
  - relative

I was angry that he destroyed **my** tent.

The Old Subtypes

- Old (6)
  - identity
  - **event**
  - general
  - generic
  - ident_generic
  - relative

They asked me to *put my phone number* on the form. **That** I think is not needed.

The Old Subtypes

- Old (6)
  - identity
  - event
  - **general**
  - generic
  - ident_generic
  - relative

Personal pronouns referring to the dialogue participants

The Old Subtypes

– Old (6)
  - identity
  - event
  - general
  - **generic**
  - ident_generic
  - relative

I think to correct the judicial system, **you** have to get the lawyer out of it

The Old Subtypes

– Old (6)
  - identity
  - event
  - general
  - generic
  - **ident_generic**
  - relative

a coreference chain of generic pronouns

The Old Subtypes

– Old (6)
  - identity
  - event
  - general
  - generic
  - ident_generic
  - **relative**

the alcoholic **that** charges up all the bills on the credit card

The Mediated Subtypes

Generally known entities, e.g., **the Earth**, **France**, …

- Mediated (9)
  - **general**
  - event
  - bound
  - part
  - situation
  - set
  - poss
  - func_value
  - aggregation

The Mediated Subtypes

– Mediated (9)
  - general
  - **event**
  - bound
  - part
  - situation
  - set
  - poss
  - func_value
  - aggregation

We were *travelling in Miami*, and **the bus** was very full.

The Mediated Subtypes

Every cat ate **its** dinner.

- Mediated (9)
  - general
  - event
  - **bound**
  - part
  - situation
  - set
  - poss
  - func_value
  - aggregation

The Mediated Subtypes

– Mediated (9)
  - general
  - event
  - bound
  - **part**
  - situation
  - set
  - poss
  - func_value
  - aggregation

He passed by Jan's *house* and saw **the door** was painted red.

The Mediated Subtypes

Mary went to *John's ranch* and saw that there were only **a few horses**.

- Mediated (9)
  - general
  - event
  - bound
  - part
  - **situation**
  - set
  - poss
  - func_value
  - aggregation

The Mediated Subtypes

What we try to do to stick to *our monthly budget* is we pretty much have **the house payment**.

– Mediated (9)
  - general
  - event
  - bound
  - part
  - situation
  - **set**
  - poss
  - func_value
  - aggregation

The Mediated Subtypes

- IS subtypes
  - Mediated (9)
    - general
    - event
    - bound
    - part
    - situation
    - set
    - poss
    - **func_value**
    - aggregation

The temperature rose to **30 degrees**.

The Mediated Subtypes

- IS subtypes
  - Mediated (9)
    - general
    - event
    - bound
    - part
    - situation
    - set
    - poss
    - func_value
    - **aggregation**

I have a son … **My son and I** like to play chess after dinner

Automatic Fine-Grained IS Determination

- is a hard problem
  - may require world knowledge and semantic understanding of a text
- but it benefits other NLP tasks such as anaphora resolution
  - NPs whose IS are new are non-anaphoric and hence should not be resolved
  - Identification of set-subset and part-whole relations is useful for bridging anaphora resolution

Related Work

- Some work on coarse-grained (i.e., 3-class) IS determination
  - Nissim (2006): 7 string-matching and grammatical features
  - Rahman and Ng (2011): augment Nissim's feature set with lexical and syntactic features
  - F-score on new entities generally very poor: ~46%

Related Work

- To our knowledge, we are the first to tackle fine-grained IS determination
- Hypothesized the poor performance on new entities can be attributed to lack of knowledge
- Propose a **knowledge-rich** approach to fine-grained IS determination

Plan for the talk

- What is Information Status?
- Fine-Grained Information Status
  - Rule-based Approach
  - Learning-based Approach
  - Evaluation
- Conclusion

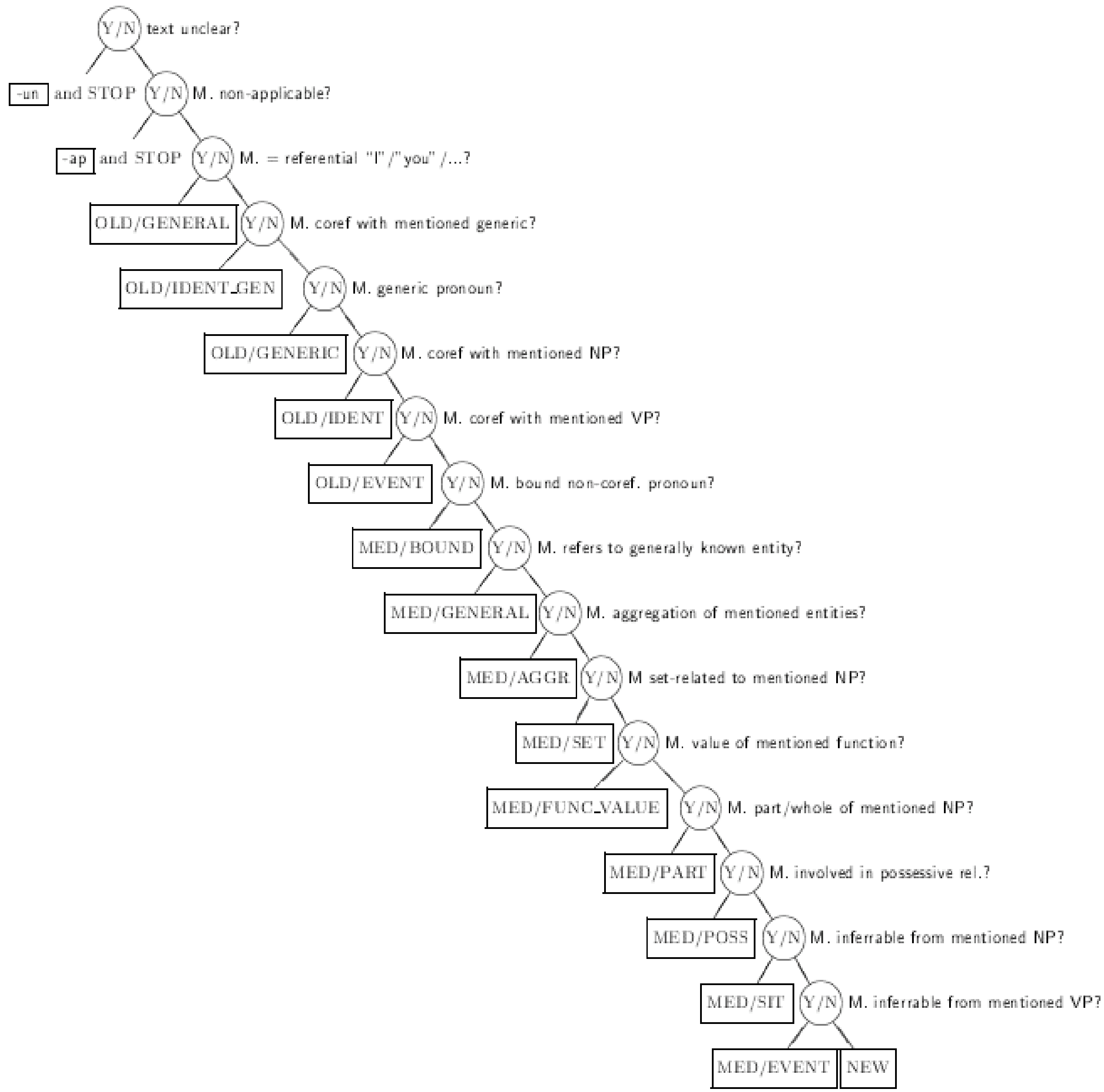
Rule-Based Approach

- 18 manually composed rules
  - roughly one rule for each IS subtype

- Models after the decision tree that Nissim et al. (2004) relied on for manually annotating IS subtypes

- Knowledge-rich approach: some rules use information extracted from WordNet, FrameNet, and ReVerb

Nissim et al.'s tree for subtype annotation



7 Rules for Old Subtypes

1. if the NP is “I” or “you” and it is not part of a coreference chain, then
   subtype := old/general

2. if the NP is “you” or “they” and it is anaphoric, then
   subtype := old/ident generic

3. if the NP is “you” or “they”, then
   subtype := old/generic

4. if the NP is “whatever” or an indefinite pronoun prefixed by “some” or “any” (e.g., “somebody”), then
   subtype := old/generic

7 Rules for Old Subtypes

5. if the NP is an anaphoric pronoun other than “that”, or its string is identical to that of a preceding NP, then
subtype := old/ident

6. if the NP is “that” and it is coreferential with the immediately preceding word, then
subtype := old/relative

7. if the NP is “it”, “this” or “that”, and it is not anaphoric, then
subtype := old/event

9 Rules for Mediated Subtypes

8. if the NP is pronominal and is not anaphoric, then

   subtype := med/bound

9. if the NP contains “and” or “or”, then

   subtype := med/aggregation

10. if the NP is a multi-word phrase that (1) begins with “so much”, “something”, “somebody”, “someone”, or (2) has “another”, “anyone”, “other”, “of” or “type” as neither its first nor last word, then

    subtype := med/set

11. if the NP contains a hyponym of the word “value” in WordNet, then

    subtype := med/func value

12. if the NP is involved in a part-whole relation with a preceding NP based on information extracted from ReVerb’s output, then

    subtype := med/part

9 Rules for Mediated Subtypes

13. if the NP is of the form "X's Y" or "poss-pro Y", where X and Y are NPs and poss-pro is a possessive pronoun, then

    subtype := med/poss

14. if the NP fills an argument of a FrameNet frame set up by a preceding NP or verb, then

    subtype := med/situation

15. if the head of the NP and one of the preceding verbs in the same sentence share the same WordNet hypernym which is not in synsets that appear one of the top five levels of the noun/verb hierarchy, then

    subtype := med/event

16. if the NP is a named entity (NE) or starts with "the", then

    subtype := med/general

2 More Rules

17. **Memorization** rule:

    if the NP appears in the training set, then

    subtype := its most frequent IS subtype in the training set

18. **Default** rule:

    subtype := new

Plan for the talk

- What is Information Status?
- Fine-Grained Information Status
  - Rule-based Approach
  - Learning-based Approach
  - Evaluation
- Conclusion

Learning-Based Approach

- Leverages the manually-crafted rules as features
- Five different feature sets
  - Rule Conditions (17)
  - Rule Predictions (17)
  - Markable Predictions (17)
  - Markables (209751)
  - Unigrams (119704)

1. Rule Conditions

- 17 binary features from the 17 hand-crafted rules (memorization rule excluded)

- Assuming that Rule i is of the form A_i -> B_i (i.e., A_i is the condition that must be satisfied in order to predict B_i as subtype), we define binary feature f_i as $\neg A_1 \wedge \neg A_2 \wedge \ldots \neg A_{i-1} \wedge A_i$

2. Rule Predictions

- 17 features based on the predictions of our 17 hand-crafted rules.
- One for each B_i
  - where in A_i -> B_i , A_i is the condition that must be satisfied in order to predict B_i as the subtype.

3. Markable Predictions

- 16 binary features, one for each IS subtype, which encode the prediction made by the memorization rule
  - Most frequent IS subtype in the training set.

4. Markables

- One binary feature for each markable appearing in the training set indicating its presence/absence
- Total 209,751 features

5. Unigrams

- One binary feature for each unigram appearing in the training set indicating its presence/absence
- Total 119,704 features.

Plan for the talk

- What is Information Status?
- Fine-Grained Information Status
  - Rule-based Approach
  - Learning-based Approach
  - Evaluation
- Conclusion

Evaluation

- Nissim et al.'s (2004) dataset
- 147 Switchboard dialogues
  - 117 for training, 30 for testing
- Total 58,835 NPs
  - We used gold-standard NPs for evaluation
- 16-class classification problem
- Train a multi-class SVM classifier on the training instances using SVM*multiclass* .
- Using both **Gold Coreference** from annotation and **Automatic Coreference** using *Stanford Deterministic Coreference System* (Lee et al., 2011).

IS Subtype	Rule-Based Approach		Learning-Based Approach	
	Gold Coref	Auto Coref	Gold Coref	Auto Coref
old/ident	77.8	58.7	84.0	69.5
old/event	66.7	53.8	92.8	4.5
old/general	82.3	77.6	95.6	90.2
old/generic	55.5	39.5	81.3	54.5
old/ident_generic	59.9	35.7	69.1	46.0
old/relative	61.3	59.0	76.7	54.4
med/general	23.8	23.6	89.4	77.7
med/bound	30.1	30.1	36.9	5.1
med/part	32.7	32.7	83.3	83.3
med/situation	44.6	44.6	79.7	80.2
med/event	18.9	18.9	63.3	63.3
med/set	70.8	67.4	89.1	87.2
med/poss	65.6	65.6	92.8	93.9
med/func_value	77.6	77.6	87.0	87.0
med/aggregation	49.9	49.6	78.6	88.6
new	57.0	56.7	87.4	86.9
ALL	**66.0**	**57.4**	**86.4**	**78.7**

Observations

- Learning-based approach beats by 20.4% (Gold coreference) and 21.3% (Auto coreference)
  - Machine learning has “transformed” a ruleset that achieves medicore performance into a system that achieves relatively high performance

- Coreference plays a crucial role in subtype classification
  - Accuracies could increase by up to 7.7-8.6% if we solely improved coreference performance

Observations (Cont')

- F-score of the new class increases by 30 points
  - Simultaneous rise in recall and precision
- Rules that rely on sophisticated knowledge (e.g., rules for med/part, med/situation, and med/event) all achieved perfect precision but low recall
  - Machine learning helps substantially improve recall

Feature Ablation Experiments

- Feature ablation : Train/test after removing each feature set separately

Feature Type	Gold Coreference	Automatic Coreference
All features	86.4	78.7
- Rule Predictions	77.5	70.0
- Rule Conditions	81.1	71.0
- Markable Predictions	72.4	64.7
- Markables	83.2	75.5
- Unigrams	74.4	58.6

- Performance drops significantly ($p < 0.05$, paired t-test) whenever a feature type is removed

Single-Feature Classifiers

- Train/test classifier using exactly one type of features

Feature Type	Gold Coreference	Automatic Coreference
Rule Predictions	49.1	45.2
Rule Conditions	58.1	28.9
Markable Predictions	39.7	39.7
Markables	10.4	10.4
Unigrams	56.8	56.8

- Markable Predictions and Unigrams are the most important feature groups

Plan for the talk

- What is Information Status?
- Fine-Grained Information Status
  - Rule-based Approach
  - Learning-based Approach
  - Evaluation
- Conclusion

Conclusion

- We proposed a rule-based approach and learning-based approach to the task of fine-grained information status determination
  - We created sophisticated features using FrameNet, WordNet, ReVerb, etc.
- Learning-based approach beats rule-based approach by around 20% in accuracy.
- Accuracy could be improved by improving the accuracy of identifying coreference chains