Joint Segmentation and Classification of Dialog Acts in Multiparty Meetings

IDIAP, July 28th 2006

Matthias Zimmermann

zimmerma@icsi.berkeley.edu

International Computer Science Institute, Berkeley, USA

Organization

1. Introduction
2. Performance Metrics
3. Hidden-Event LM
4. Maximum Entropy
5. Conditional Random Fields
6. Conclusions and Outlook

Acknowledgments

Andreas Stolcke, Liz Shriberg, Yang Liu

Introduction

Segmentation of a meeting into its Dialog Acts (DAs), where

Introduction

Segmentation of a meeting into its Dialog Acts (DAs), where

Meeting Transcription per audio channel (participant)

Introduction

Segmentation of a meeting into its Dialog Acts (DAs), where

Meeting	Transcription per audio channel (participant)
DAs	Fundamental building blocks of conversations can be viewed as typed utterances

Introduction

Segmentation of a meeting into its Dialog Acts (DAs), where

Meeting	Transcription per audio channel (participant)
DAs	Fundamental building blocks of conversations can be viewed as typed utterances

DA types (for this talk):

Introduction

Segmentation of a meeting into its Dialog Acts (DAs), where

Meeting	Transcription per audio channel (participant)
DAs	Fundamental building blocks of conversations can be viewed as typed utterances

DA types (for this talk):

Statements	yeah that's right. maybe.

Introduction

Segmentation of a meeting into its Dialog Acts (DAs), where

Meeting	Transcription per audio channel (participant)
DAs	Fundamental building blocks of conversations can be viewed as typed utterances

DA types (for this talk):

Statements	yeah that's right.	maybe.
Questions	right?	why twenty eight?

Introduction

Segmentation of a meeting into its Dialog Acts (DAs), where

Meeting	Transcription per audio channel (participant)
DAs	Fundamental building blocks of conversations can be viewed as typed utterances

DA types (for this talk):

Statements	yeah that's right. maybe.
Questions	right? why twenty eight?
Backchannels	yeah right

Introduction

Segmentation of a meeting into its Dialog Acts (DAs), where

Meeting	Transcription per audio channel (participant)
DAs	Fundamental building blocks of conversations can be viewed as typed utterances

DA types (for this talk):

Statements	yeah that's right. maybe.
Questions	right? why twenty eight?
Backchannels	yeah right
Floorgrabbers	um- so- uh uh-

Introduction

Segmentation of a meeting into its Dialog Acts (DAs), where

Meeting	Transcription per audio channel (participant)
DAs	Fundamental building blocks of conversations can be viewed as typed utterances

DA types (for this talk):

Statements	yeah that's right. maybe.
Questions	right? why twenty eight?
Backchannels	yeah right
Floorgrabbers	um- so- uh uh-
Disruptions	but- and it doesn't-

Introduction

What do we need to do?

Introduction

What do we need to do?

Find DA segments in stream of words

Introduction

What do we need to do?

- Find DA segments in stream of words
- Find DA types for DA segments

Introduction

What do we need to do?

- Find DA segments in stream of words
- Find DA types for DA segments
- Evaluate system output against the gold standard

Introduction

What do we need to do?

- Find DA segments in stream of words
- Find DA types for DA segments
- Evaluate system output against the gold standard

Example

Input: um that's interesting what yeah i i i see

Introduction

What do we need to do?

- Find DA segments in stream of words
- Find DA types for DA segments
- Evaluate system output against the gold standard

Example

Input: um that's interesting what yeah i i i see

Gold: um- | that's interesting. | what? | yeah i i- | i see.

Introduction

What do we need to do?

- Find DA segments in stream of words
- Find DA types for DA segments
- Evaluate system output against the gold standard

Example

Input: um that's interesting what yeah i i i see

Gold:	um-	that's interesting.	what?	yeah i i-	i see.
System:	um-	that's interesting.	what?	yeah.	i i i see.

Introduction

Where do Dialog Acts come from?

Introduction

Where do Dialog Acts come from?

- Early philosophy of language
  - Most attention to sentences that state facts/actions
  - "Sam smokes habitually"

Introduction

Where do Dialog Acts come from?

- Early philosophy of language
  - Most attention to sentences that state facts/actions
  - "Sam smokes habitually"
- 1962: J.L. Austin, "How to Do Things With Words"
  - Illocutonary acts (acts in saying something)
  - "I name this ship the *Queen Elizabeth*"

Introduction

Where do Dialog Acts come from?

- Early philosophy of language
  - Most attention to sentences that state facts/actions
  - "Sam smokes habitually"
- 1962: J.L. Austin, "How to Do Things With Words"
  - Illocutonary acts (acts in saying something)
  - "I name this ship the *Queen Elizabeth*"
- 1969: J. Searle, "Speech Acts: An Essay in Philosopy of Language"
  - Investigates different illocutonary forces
  - "Does Sam smoke habitually?" (Question)
  - "Sam, smoke habitually!" (Command)
  - "Would that Sam smoked habitually." (Desire)

Introduction

What are Dialog Acts good for?

Introduction

What are Dialog Acts good for?

Natural Language Processing Techniques

- POS tagging / parsing
- noun phrase identification
- Named entity extraction
- Coreference resolution

Introduction

What are Dialog Acts good for?

- Natural Language Processing Techniques
  - POS tagging / parsing
  - noun phrase identification
  - Named entity extraction
  - Coreference resolution
- Dialog Systems
  - Question detection
  - Backchannel detection

Introduction

What are Dialog Acts good for?

- Natural Language Processing Techniques
  - POS tagging / parsing
  - noun phrase identification
  - Named entity extraction
  - Coreference resolution
- Dialog Systems
  - Question detection
  - Backchannel detection
- Interaction modeling
  - Dominance detection

Introduction

What are Dialog Acts good for?

- Natural Language Processing Techniques
  - POS tagging / parsing
  - noun phrase identification
  - Named entity extraction
  - Coreference resolution
- Dialog Systems
  - Question detection
  - Backchannel detection
- Interaction modeling
  - Dominance detection
- ...?

Introduction

Previous work on dialog acts

- Concentrated on either segmentation or classification

Introduction

Previous work on dialog acts

- Concentrated on either segmentation or classification
- However, fully automatic systems require solutions to both

Introduction

Previous work on dialog acts

- Concentrated on either segmentation or classification
- However, fully automatic systems require solutions to both

1997: V. Warnke et al. "Integrated Dialog Act Segmentation And Classification Using Prosodic Features And Language Models"

- VERBOMIL

Introduction

Previous work on dialog acts

- Concentrated on either segmentation or classification
- However, fully automatic systems require solutions to both

1997: V. Warnke et al. "Integrated Dialog Act Segmentation And Classification Using Prosodic Features And Language Models"

- VERBOMIL

2005: J. Ang et al. "Automatic Dialog Act Segmentation and Classification in Multiparty Meetings"

- ICSI meeting corpus, sequential approach

Introduction

Previous work on dialog acts

- Concentrated on either segmentation or classification
- However, fully automatic systems require solutions to both

1997: V. Warnke et al. "Integrated Dialog Act Segmentation And Classification Using Prosodic Features And Language Models"

- VERBOMIL

2005: J. Ang et al. "Automatic Dialog Act Segmentation and Classification in Multiparty Meetings"

- ICSI meeting corpus, sequential approach
- Segmentation: hidden event LM (words), decision trees (pauses)

Introduction

Previous work on dialog acts

- Concentrated on either segmentation or classification
- However, fully automatic systems require solutions to both

1997: V. Warnke et al. "Integrated Dialog Act Segmentation And Classification Using Prosodic Features And Language Models"

- VERBOMIL

2005: J. Ang et al. "Automatic Dialog Act Segmentation and Classification in Multiparty Meetings"

- ICSI meeting corpus, sequential approach
- Segmentation: hidden event LM (words), decision trees (pauses)
- Classification: maximum entropy (words)

Introduction

This work: Joint task as standard classification problem

Introduction

This work: Joint task as standard classification problem

- Classification of interword boundaries

Introduction

This work: Joint task as standard classification problem

- Classification of interword boundaries
- n DA types, $n + 1$ classes

Introduction

This work: Joint task as standard classification problem

- Classification of interword boundaries
- n DA types, $n + 1$ classes
- Any classification method may be used

Introduction

This work: Joint task as standard classification problem

- Classification of interword boundaries
- n DA types, $n + 1$ classes
- Any classification method may be used

Mapping procedure

Introduction

This work: Joint task as standard classification problem

- Classification of interword boundaries
- n DA types, $n + 1$ classes
- Any classification method may be used

Mapping procedure

Gold ... for- | okay. | oh. | this friday? | yeah.

Introduction

This work: Joint task as standard classification problem

- Classification of interword boundaries
- n DA types, $n + 1$ classes
- Any classification method may be used

Mapping procedure

Gold	... for- okay. oh. this friday? yeah.
Gold mapped	... for D| okay S| oh S| this * friday Q| yeah S|

Introduction

This work: Joint task as standard classification problem

- Classification of interword boundaries
- n DA types, $n + 1$ classes
- Any classification method may be used

Mapping procedure

Gold	... for- okay. oh. this friday? yeah.
Gold mapped	... for D| okay S| oh S| this * friday Q| yeah S|
System output	... for S| okay S| oh * this * friday S| yeah B|

Introduction

This work: Joint task as standard classification problem

- Classification of interword boundaries
- n DA types, $n + 1$ classes
- Any classification method may be used

Mapping procedure

Gold	... for- okay. oh. this friday? yeah.
Gold mapped	... for D| okay S| oh S| this * friday Q| yeah S|
System output	... for S| okay S| oh * this * friday S| yeah B|
System	... for. okay. oh this friday. yeah.

Performance Metrics

Why should we care?

Performance Metrics

Why should we care?

- Metrics are fundamental for tuning and evaluating systems

Performance Metrics

Why should we care?

- Metrics are fundamental for tuning and evaluating systems
- Only few metrics exist that address the joint problem

Performance Metrics

Why should we care?

- Metrics are fundamental for tuning and evaluating systems
- Only few metrics exist that address the joint problem
- No metrics that are directly based on DAs

Performance Metrics

Why should we care?

- Metrics are fundamental for tuning and evaluating systems
- Only few metrics exist that address the joint problem
- No metrics that are directly based on DAs

Previously proposed Metrics

- Accuracy, DA labels only (Warnke et al.)

Performance Metrics

Why should we care?

- Metrics are fundamental for tuning and evaluating systems
- Only few metrics exist that address the joint problem
- No metrics that are directly based on DAs

Previously proposed Metrics

- Accuracy, DA labels only (Warnke et al.)
- NIST, boundary based metric

Performance Metrics

Why should we care?

- Metrics are fundamental for tuning and evaluating systems
- Only few metrics exist that address the joint problem
- No metrics that are directly based on DAs

Previously proposed Metrics

- Accuracy, DA labels only (Warnke et al.)
- NIST, boundary based metric
- Lenient, word labels only (Ang et al.)
- Strict, segment boundaries and word labels (Ang et al.)

Performance Metrics

Gold	`S|Q.Q.Q.Q|S.S.S|B|S.S|`
System	`S|Q|S|Q.Q|D.D.D|S.S|S|`

Performance Metrics

Gold	S|Q.Q.Q.Q|S.S.S|B|S.S|
System	S|Q|S|Q.Q|D.D.D|S.S|S|
NIST	c f f c s m f c

Performance Metrics

Gold	S|Q.Q.Q.Q|S.S.S|B|S.S|
System	S|Q|S|Q.Q|D.D.D|S.S|S|
NIST	c f f c s m f c
Strict	c|e e e e|e e e|e|e e|

Performance Metrics

Gold	S|Q.Q.Q.Q|S.S.S|B|S.S|
System	S|Q|S|Q.Q|D.D.D|S.S|S|
NIST	c f f c s m f c
Strict	c|e e e e|e e e|e|e e|
Recall	c| e | e |e| e |

Performance Metrics

Gold	S|Q.Q.Q.Q|S.S.S|B|S.S|
System	S|Q|S|Q.Q|D.D.D|S.S|S|
NIST	c f f c s m f c
Strict	c|e e e e|e e e|e|e e|
Recall	c| e | e |e| e |
Precision	c|e|e| e | e | e |e|

Performance Metrics

Gold	S|Q.Q.Q.Q|S.S.S|B|S.S|
System	S|Q|S|Q.Q|D.D.D|S.S|S|
NIST	c f f c s m f c
Strict	c|e e e e|e e e|e|e e|
Recall	c| e | e |e| e |
Precision	c|e|e| e | e | e |e|

Metric	Counts	Reference	Rate
NIST	3 FA, 1 miss, 1 subst.	5 boundaries	100%
Strict	10 bad word labels	11 words	91%

Performance Metrics

Gold	S|Q.Q.Q.Q|S.S.S|B|S.S|
System	S|Q|S|Q.Q|D.D.D|S.S|S|
NIST	c f f c s m f c
Strict	c|e e e e|e e e|e|e e|
Recall	c| e | e |e| e |
Precision	c|e|e| e | e | e |e|

Metric	Counts	Reference	Rate
NIST	3 FA, 1 miss, 1 subst.	5 boundaries	100%
Strict	10 bad word labels	11 words	91%
Recall	1 correct dialog act	5 dialog acts	20%
Precision	1 correct dialog act	7 dialog acts	14%
F-Measure	-	-	17%

Hidden Event Language Model

Hidden Event LM (HELM)

- Standard n-gram modeling for stream of words and hidden events

Hidden Event Language Model

Hidden Event LM (HELM)

- Standard n-gram modeling for stream of words and hidden events
- Here: Typed DA boundaries act as hidden events

Hidden Event Language Model

Hidden Event LM (HELM)

- Standard n-gram modeling for stream of words and hidden events
- Here: Typed DA boundaries act as hidden events

Training: Hidden events treated as normal words

Hidden Event Language Model

Hidden Event LM (HELM)

- Standard n-gram modeling for stream of words and hidden events
- Here: Typed DA boundaries act as hidden events

Training: Hidden events treated as normal words

well D uh that's pretty good i think S yeah B

Hidden Event Language Model

Hidden Event LM (HELM)

- Standard n-gram modeling for stream of words and hidden events
- Here: Typed DA boundaries act as hidden events

Training: Hidden events treated as normal words

well D uh that's pretty good i think S yeah B

Testing: Insert event according to decision rule

Hidden Event Language Model

Hidden Event LM (HELM)

- Standard n-gram modeling for stream of words and hidden events
- Here: Typed DA boundaries act as hidden events

Training: Hidden events treated as normal words

well D uh that's pretty good i think S yeah B

Testing: Insert event according to decision rule

well uh that's pretty good i think yeah thanks

Hidden Event Language Model

Hidden Event LM (HELM)

- Standard n-gram modeling for stream of words and hidden events
- Here: Typed DA boundaries act as hidden events

Training: Hidden events treated as normal words

well D uh that's pretty good i think S yeah B

Testing: Insert event according to decision rule

well uh that's pretty good i think yeah thanks

$p($ * $|$ pretty $)$, $p($S $|$ pretty $)$, $p($Q $|$ pretty $)$, . . . $p($ D $|$ pretty $)$

Hidden Event Language Model

Hidden Event LM (HELM)

- Standard n-gram modeling for stream of words and hidden events
- Here: Typed DA boundaries act as hidden events

Training: Hidden events treated as normal words

well D uh that's pretty good i think S yeah B

Testing: Insert event according to decision rule

well uh that's pretty good i think yeah thanks

$p($ * | pretty $)$, $p($S | pretty $)$, $p($Q | pretty $)$, . . . $p($ D | pretty $)$

Decision rule

IF $p($ * $| \ldots) > \theta$ THEN non-event

ELSE insert DA type associated with highest posterior

Hidden Event Language Model

Experimental Setup (Ang et al. 2005)

- ICSI meeting corpus with DA annotations (MRDA)
- 51 meetings for training, 11 for validation, and 11 for testing

Hidden Event Language Model

Experimental Setup (Ang et al. 2005)

- ICSI meeting corpus with DA annotations (MRDA)
- 51 meetings for training, 11 for validation, and 11 for testing
- 2 conditions: Reference text, and STT* output

*: Average WER: 39%

Hidden Event Language Model

Experimental Setup (Ang et al. 2005)

- ICSI meeting corpus with DA annotations (MRDA)
- 51 meetings for training, 11 for validation, and 11 for testing
- 2 conditions: Reference text, and STT* output
- 5 DA types†

*: Average WER: 39%

†: Backchannel, Disruption, Floorgrabber, Question, Statement

Hidden Event Language Model

Experimental Results

- System based on HELM
- Comparison with sequential solution (Ang et al. 2005)

Condition	System	Strict	Recall
	Ang et al.	64.4	45.6
Ref	Ang et al.*	72.4	35.9
STT			

*: Ang et al, pause duration not used

Hidden Event Language Model

Experimental Results

- System based on HELM
- Comparison with sequential solution (Ang et al. 2005)

Condition	System	Strict	Recall
	Ang et al.	64.4	45.6
Ref	Ang et al.*	72.4	35.9
	HELM	74.3	33.5
STT			

*: Ang et al, pause duration not used

Hidden Event Language Model

Experimental Results

- System based on HELM
- Comparison with sequential solution (Ang et al. 2005)

Condition	System	Strict	Recall
Ref	Ang et al.	64.4	45.6
	Ang et al.*	72.4	35.9
	HELM	74.3	33.5
STT	Ang et al.	75.4	35.7
	Ang et al.*	82.9	26.8

*: Ang et al, pause duration not used

Hidden Event Language Model

Experimental Results

- System based on HELM
- Comparison with sequential solution (Ang et al. 2005)

Condition	System	Strict	Recall
Ref	Ang et al.	64.4	45.6
	Ang et al.*	72.4	35.9
	HELM	74.3	33.5
STT	Ang et al.	75.4	35.7
	Ang et al.*	82.9	26.8
	HELM	83.8	26.1

*: Ang et al, pause duration not used

Hidden Event Language Model

Findings

- HELM based approach on words alone

Hidden Event Language Model

Findings

- HELM based approach on words alone
- Baseline results for joint segmentation and classification

Hidden Event Language Model

Findings

- HELM based approach on words alone
- Baseline results for joint segmentation and classification
- OK results for simplicity of approach

Hidden Event Language Model

Findings

- HELM based approach on words alone
- Baseline results for joint segmentation and classification
- OK results for simplicity of approach
- Prosody should be included

Maximum Entropy

Maximum Entropy (MaxEnt)

- Old principle, introduced in NLP 1996 by A. Berger et al. "A Maximum Entropy Approach to Natural Language Processing"

Maximum Entropy

Maximum Entropy (MaxEnt)

- Old principle, introduced in NLP 1996 by A. Berger et al. "A Maximum Entropy Approach to Natural Language Processing"
- 2002, J. Huang and G. Zweig "Maximum Entropy Model for Punctuation Annotation from Speech"

Maximum Entropy

Maximum Entropy (MaxEnt)

- Old principle, introduced in NLP 1996 by A. Berger et al. "A Maximum Entropy Approach to Natural Language Processing"
- 2002, J. Huang and G. Zweig "Maximum Entropy Model for Punctuation Annotation from Speech"
- Words and pause durations are integrated in a single framework

Maximum Entropy

Maximum Entropy (MaxEnt)

- Old principle, introduced in NLP 1996 by A. Berger et al. "A Maximum Entropy Approach to Natural Language Processing"
- 2002, J. Huang and G. Zweig "Maximum Entropy Model for Punctuation Annotation from Speech"
- Words and pause durations are integrated in a single framework
- Adapt this scheme to the 6-way DA task

Maximum Entropy

Maximum Entropy (MaxEnt)

- Old principle, introduced in NLP 1996 by A. Berger et al. "A Maximum Entropy Approach to Natural Language Processing"
- 2002, J. Huang and G. Zweig "Maximum Entropy Model for Punctuation Annotation from Speech"
- Words and pause durations are integrated in a single framework
- Adapt this scheme to the 6-way DA task
- Combination of MaxEnt and HELM: 2000, E. Shriberg et al. "Prosody-based automatic Segmentation of Speech into Sentences and Topics"

Maximum Entropy

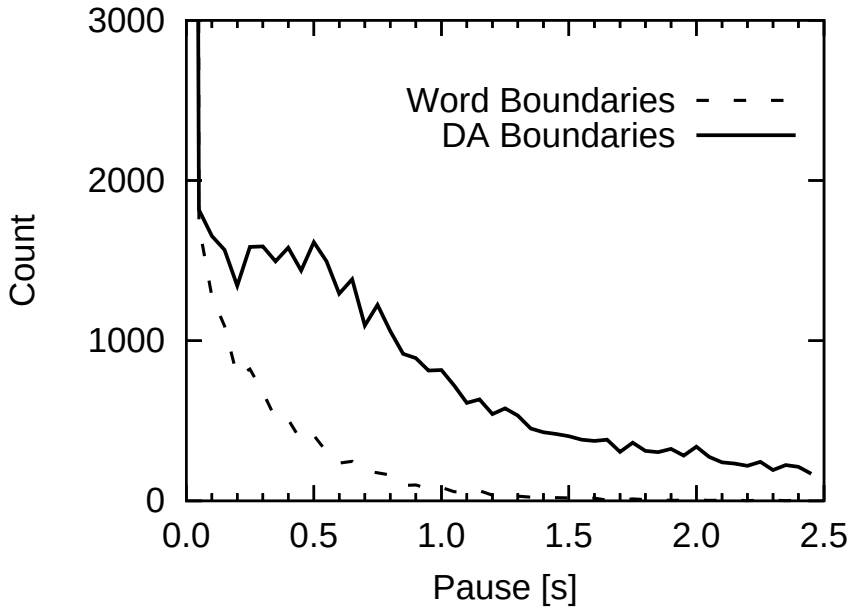
Pause Durations

- Pause durations are the single most helpful prosodic feature

Maximum Entropy

Pause Durations

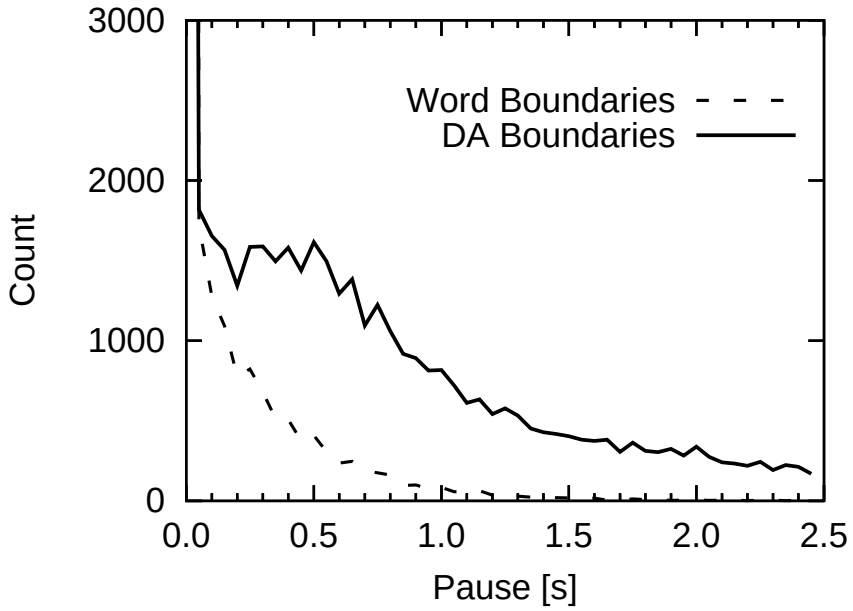
- Pause durations are the single most helpful prosodic feature
- Different distributions for DA boundaries and non-boundaries



Maximum Entropy

Pause Durations

- Pause durations are the single most helpful prosodic feature
- Different distributions for DA boundaries and non-boundaries
- Pause durations and neighbouring words are correlated



Word Pair	Pause [s]	Count
you know	0.006	2876
the the	0.176	988
um um	0.935	55
because um	0.399	34
um yeah	0.746	30

Maximum Entropy

Feature Extraction

- Binary indicator functions as features

Maximum Entropy

Feature Extraction

- Binary indicator functions as features
- Discretize pause duration into histogram bins

Maximum Entropy

Feature Extraction

- Binary indicator functions as features
- Discretize pause duration into histogram bins

Example

uh	that's	pretty	p=0	good	i
w_{i-2}	w_{i-1}	w_i	p_i	w_{i+1}	w_{i+2}

Maximum Entropy

Feature Extraction

- Binary indicator functions as features
- Discretize pause duration into histogram bins

Example

uh	that's	pretty	p=0	good	i
w_{i-2}	w_{i-1}	w_i	p_i	w_{i+1}	w_{i+2}

Words	Feature Vector
0	p_i

Maximum Entropy

Feature Extraction

- Binary indicator functions as features
- Discretize pause duration into histogram bins

Example

uh	that's	pretty	p=0	good	i
w_{i-2}	w_{i-1}	w_i	p_i	w_{i+1}	w_{i+2}

Words	Feature Vector
0	p_i
1	p_i, w_i, (w_i, p_i)

Maximum Entropy

Feature Extraction

- Binary indicator functions as features
- Discretize pause duration into histogram bins

Example

uh	that's	pretty	p=0	good	i
w_{i-2}	w_{i-1}	w_i	p_i	w_{i+1}	w_{i+2}

Words	Feature Vector
0	p_i
1	p_i, w_i, (w_i, p_i)
2	p_i, w_i, (w_i, p_i), w_{i+1}, (p_i, w_{i+1}), (w_i, w_{i+1})

Maximum Entropy

Feature Extraction

- Binary indicator functions as features
- Discretize pause duration into histogram bins

Example

uh	that's	pretty	p=0	good	i
w_{i-2}	w_{i-1}	w_i	p_i	w_{i+1}	w_{i+2}

Words	Feature Vector
0	p_i
1	p_i, w_i, (w_i, p_i)
2	p_i, w_i, (w_i, p_i), w_{i+1}, (p_i, w_{i+1}), (w_i, w_{i+1})
3	p_i, w_i, (w_i, p_i), w_{i+1}, (p_i, w_{i+1}), (w_i, w_{i+1}), w_{i-1}, (w_{i-1}, w_i)

Maximum Entropy

Feature Extraction

- Binary indicator functions as features
- Discretize pause duration into histogram bins

Example

uh	that's	pretty	p=0	good	i
w_{i-2}	w_{i-1}	w_i	p_i	w_{i+1}	w_{i+2}

Words	Feature Vector
0	p_i
1	p_i, w_i, (w_i, p_i)
2	p_i, w_i, (w_i, p_i), w_{i+1}, (p_i, w_{i+1}), (w_i, w_{i+1})
3	p_i, w_i, (w_i, p_i), w_{i+1}, (p_i, w_{i+1}), (w_i, w_{i+1}), w_{i-1}, (w_{i-1}, w_i)
4	$\ldots$ w_{i+2}, (w_{i+1}, w_{i+2})

Maximum Entropy

Experimental Results

- Identical setup as before
- Combination of MaxEnt and HELM
- Compare with sequential solution (Ang et al. 2005)

Condition	System	Strict	Recall
	Ang et al.	64.4	45.6
Ref			
STT			

Maximum Entropy

Experimental Results

- Identical setup as before
- Combination of MaxEnt and HELM
- Compare with sequential solution (Ang et al. 2005)

Condition	System	Strict	Recall
	Ang et al.	64.4	45.6
Ref	No word context	65.0	44.1
STT			

Maximum Entropy

Experimental Results

- Identical setup as before
- Combination of MaxEnt and HELM
- Compare with sequential solution (Ang et al. 2005)

Condition	System	Strict	Recall
	Ang et al.	64.4	45.6
Ref	No word context	65.0	44.1
	With 4 word context	62.8	49.0
STT			

Maximum Entropy

Experimental Results

- Identical setup as before
- Combination of MaxEnt and HELM
- Compare with sequential solution (Ang et al. 2005)

Condition	System	Strict	Recall
Ref	Ang et al.	64.4	45.6
	No word context	65.0	44.1
	With 4 word context	62.8	49.0
STT	Ang et al.	75.4	35.7

Maximum Entropy

Experimental Results

- Identical setup as before
- Combination of MaxEnt and HELM
- Compare with sequential solution (Ang et al. 2005)

Condition	System	Strict	Recall
Ref	Ang et al.	64.4	45.6
	No word context	65.0	44.1
	With 4 word context	62.8	49.0
STT	Ang et al.	75.4	35.7
	No word context	76.2	34.9

Maximum Entropy

Experimental Results

- Identical setup as before
- Combination of MaxEnt and HELM
- Compare with sequential solution (Ang et al. 2005)

Condition	System	Strict	Recall
Ref	Ang et al.	64.4	45.6
	No word context	65.0	44.1
	With 4 word context	62.8	49.0
STT	Ang et al.	75.4	35.7
	No word context	76.2	34.9
	With 4 word context	73.6	37.4

Maximum Entropy

Findings

- Combination of HELM and MaxEnt

Maximum Entropy

Findings

- Combination of HELM and MaxEnt
- Joint approach outperforms the sequential approach

Maximum Entropy

Findings

- Combination of HELM and MaxEnt
- Joint approach outperforms the sequential approach
- STT conditions: MaxEnt alone comparable Ang et al.

Maximum Entropy

Findings

- Combination of HELM and MaxEnt
- Joint approach outperforms the sequential approach
- STT conditions: MaxEnt alone comparable Ang et al.
- Joint modeling of words and pauses essential

Maximum Entropy

Findings

- Combination of HELM and MaxEnt
- Joint approach outperforms the sequential approach
- STT conditions: MaxEnt alone comparable Ang et al.
- Joint modeling of words and pauses essential
- Modeling of DA sequence should be included

Conditional Random Fields

Conditional Random Fields (CRF)

- Introduced in 2001 by J. Lafferty et al. "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data"

Conditional Random Fields

Conditional Random Fields (CRF)

- Introduced in 2001 by J. Lafferty et al. "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data"
- Classification optimizes sequence not single instances

Conditional Random Fields

Conditional Random Fields (CRF)

- Introduced in 2001 by J. Lafferty et al. "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data"
- Classification optimizes sequence not single instances
- Allows for implicit representation of a DA grammar (DA type bigram)

Conditional Random Fields

Conditional Random Fields (CRF)

- Introduced in 2001 by J. Lafferty et al. "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data"
- Classification optimizes sequence not single instances
- Allows for implicit representation of a DA grammar (DA type bigram)
- Works out of the box for the MaxEnt feature set

Conditional Random Fields

Labeling

- Use previous interword boundary labels (6-way classification)

Conditional Random Fields

Labeling

- Use previous interword boundary labels (6-way classification)
- Unfortunately, this is sub-optimal for longer DAs

Conditional Random Fields

Labeling

- Use previous interword boundary labels (6-way classification)
- Unfortunately, this is sub-optimal for longer DAs
- Also try to preserve DA type over longer DAs

Conditional Random Fields

Labeling

- Use previous interword boundary labels (6-way classification)
- Unfortunately, this is sub-optimal for longer DAs
- Also try to preserve DA type over longer DAs

Example

oh okay what do you think anything

Conditional Random Fields

Labeling

- Use previous interword boundary labels (6-way classification)
- Unfortunately, this is sub-optimal for longer DAs
- Also try to preserve DA type over longer DAs

Example

	oh	okay	what	do	you	think	anything
6 Labels	*	S|	*	*	*	Q|	S|

Conditional Random Fields

Labeling

- Use previous interword boundary labels (6-way classification)
- Unfortunately, this is sub-optimal for longer DAs
- Also try to preserve DA type over longer DAs

Example

	oh	okay	what	do	you	think	anything
6 Labels	*	S|	*	*	*	Q|	S|
10 Labels	s	S|	q	q	q	Q|	S|

Conditional Random Fields

Labeling

- Use previous interword boundary labels (6-way classification)
- Unfortunately, this is sub-optimal for longer DAs
- Also try to preserve DA type over longer DAs

Example

	oh	okay	what	do	you	think	anything
6 Labels	*	S|	*	*	*	Q|	S|
10 Labels	s	S|	q	q	q	Q|	S|

- Potential drawback: 10-way classification

Conditional Random Fields

Experimental Results

- Setup as before, but lower WER (35.4% instead of 39%)
- CRF with 6-way and 10-way classification
- Comparison with sequential solution (Ang et al. 2005)

Condition	System	Strict	F-Measure
	Ang et al.	65.3	48.5
Ref			
STT			

Conditional Random Fields

Experimental Results

- Setup as before, but lower WER (35.4% instead of 39%)
- CRF with 6-way and 10-way classification
- Comparison with sequential solution (Ang et al. 2005)

Condition	System	Strict	F-Measure
	Ang et al.	65.3	48.5
Ref	CRF 6-way	61.6	52.5
STT			

Conditional Random Fields

Experimental Results

- Setup as before, but lower WER (35.4% instead of 39%)
- CRF with 6-way and 10-way classification
- Comparison with sequential solution (Ang et al. 2005)

Condition	System	Strict	F-Measure
Ref	Ang et al.	65.3	48.5
	CRF 6-way	61.6	52.5
	CRF 10-way	60.2	53.4
STT			

Conditional Random Fields

Experimental Results

- Setup as before, but lower WER (35.4% instead of 39%)
- CRF with 6-way and 10-way classification
- Comparison with sequential solution (Ang et al. 2005)

Condition	System	Strict	F-Measure
Ref	Ang et al.	65.3	48.5
	CRF 6-way	61.6	52.5
	CRF 10-way	60.2	53.4
STT	Ang et al.	74.6	39.5

Conditional Random Fields

Experimental Results

- Setup as before, but lower WER (35.4% instead of 39%)
- CRF with 6-way and 10-way classification
- Comparison with sequential solution (Ang et al. 2005)

Condition	System	Strict	F-Measure
Ref	Ang et al.	65.3	48.5
	CRF 6-way	61.6	52.5
	CRF 10-way	60.2	53.4
STT	Ang et al.	74.6	39.5
	CRF 6-way	70.2	43.7

Conditional Random Fields

Experimental Results

- Setup as before, but lower WER (35.4% instead of 39%)
- CRF with 6-way and 10-way classification
- Comparison with sequential solution (Ang et al. 2005)

Condition	System	Strict	F-Measure
Ref	Ang et al.	65.3	48.5
	CRF 6-way	61.6	52.5
	CRF 10-way	60.2	53.4
STT	Ang et al.	74.6	39.5
	CRF 6-way	70.2	43.7
	CRF 10-way	70.3	44.1

Conditional Random Fields

Confusion Matrix (6-way classification)

Gold	B	D	F	Q	S	DAs	F-Measure
B	1500	5	15	7	329	1960	82.8
D	55	636	60	14	334	2237	37.4
F	64	83	871	4	238	1924	52.1
Q	27	18	2	259	382	1164	34.1
S	410	63	53	33	4528	8918	50.8

Conditional Random Fields

Confusion Matrix (6-way classification)

Gold	B	D	F	Q	S	DAs	F-Measure
B	1500	5	15	7	329	1960	82.8
D	55	636	60	14	334	2237	37.4
F	64	83	871	4	238	1924	52.1
Q	27	18	2	259	382	1164	34.1
S	410	63	53	33	4528	8918	50.8

Confusion Matrix (10-way classification)

Gold	B	D	F	Q	S	DAs	F-Measure
B	1526	2	16	7	308	1960	84.1
D	59	662	63	33	294	2237	38.6
F	66	75	912	7	251	1924	54.2
Q	26	27	.	410	247	1164	45.2
S	431	69	55	51	4453	8918	50.6

Conditional Random Fields

Findings

- Best results so far

Conditional Random Fields

Findings

- Best results so far
- Simplest approach so far

Conditional Random Fields

Findings

- Best results so far
- Simplest approach so far
- Specifically: Boost for question detection

Conditional Random Fields

Findings

- Best results so far
- Simplest approach so far
- Specifically: Boost for question detection
- Disruptions very hard to recognize

Conclusions

Main Conclusions

Conclusions

Main Conclusions

- Investigated joint segmentation and classification of DAs

Conclusions

Main Conclusions

- Investigated joint segmentation and classification of DAs
- Joint approach substantially outperforms the sequential solution

Conclusions

Main Conclusions

- Investigated joint segmentation and classification of DAs
- Joint approach substantially outperforms the sequential solution
- The simplest systems (CRF) performed best

Conclusions

Main Conclusions

- Investigated joint segmentation and classification of DAs
- Joint approach substantially outperforms the sequential solution
- The simplest systems (CRF) performed best
- But, we just scratched at the surface of the joint task

Outlook

Future Directions

Outlook

Future Directions

- Use of different class maps (more than 5 DA types)

Outlook

Future Directions

- Use of different class maps (more than 5 DA types)
- Work with other databases (Switchboard, AMI)

Outlook

Future Directions

- Use of different class maps (more than 5 DA types)
- Work with other databases (Switchboard, AMI)
- Use many more prosodic features

Outlook

Future Directions

- Use of different class maps (more than 5 DA types)
- Work with other databases (Switchboard, AMI)
- Use many more prosodic features
- Integrate optical cues, e.g. gazing

Outlook

Future Directions

- Use of different class maps (more than 5 DA types)
- Work with other databases (Switchboard, AMI)
- Use many more prosodic features
- Integrate optical cues, e.g. gazing
- Take into account more than a single channel

Outlook

Future Directions

- Use of different class maps (more than 5 DA types)
- Work with other databases (Switchboard, AMI)
- Use many more prosodic features
- Integrate optical cues, e.g. gazing
- Take into account more than a single channel
- Consider diarization as part of the task

Outlook

Future Directions

- Use of different class maps (more than 5 DA types)
- Work with other databases (Switchboard, AMI)
- Use many more prosodic features
- Integrate optical cues, e.g. gazing
- Take into account more than a single channel
- Consider diarization as part of the task
- Build applications on top of DA recognition

Thank You