Recovering Participant Identities in Meetings from a Probabilistic Description of Vocal Interaction

Kornel Laskowski & Tanja Schultz

Cognitive Systems Lab, Universität Karlsruhe
Language Technologies Institute, Carnegie Mellon University

23 September, 2008
Outline

1. Introduction
   - Definitions
   - Motivation
   - Related Work

2. Some Concepts
   - Joint vs Independent Classification
   - *Shuffling* vs *Drawing & Shuffling* Participants
   - Features
   - Models

3. Experiments

4. Conclusions
vocal activity patterns for all $K$ participants, **seen together**

- only talkspurt start/end times = text-independence
- formally, at time $t$:

- we’ll use a discretized version
Vocal Interaction (Dabbs & Ruback, 1987)

vocal activity patterns for all $K$ participants, \textit{seen together}

- only talkspurt start/end times = text-independence
- formally, at time $t$:

we'll use a discretized version
Vocal Interaction (Dabbs & Ruback, 1987)

- vocal activity patterns for all $K$ participants, seen together
- only talkspurt start/end times = text-independence
- formally, at time $t$:
  - vocal activity of participant $k$: $q[k] \in \mathbb{N} = \{0, 1\}$
  - entire $K$-participant conversation: $q \in 
    \mathbb{N}^K$
- we’ll use a discretized version
Vocal Interaction (Dabbs & Ruback, 1987)

- vocal activity patterns for all $K$ participants, seen together
- only talkspurt start/end times = text-independence
- formally, at time $t$:
  - vocal activity of participant $k$: $q_t[k] \in V \equiv \{\square, \blacksquare\} \equiv \{0, 1\}$
  - entire $K$-participant conversation: $q_t \in V^K$
- we'll use a discretized version
Vocal Interaction (Dabbs & Ruback, 1987)

- vocal activity patterns for all $K$ participants, **seen together**
- only talkspurt start/end times = text-independence
- formally, at time $t$:
  - vocal activity of participant $k$: $q_t[k] \in \mathbb{V} \equiv \{\square, \blacksquare\} \equiv \{0, 1\}$
  - entire $K$-participant conversation: $q_t \in \mathbb{V}^K$
- we'll use a discretized version
Vocal Interaction (Dabbs & Ruback, 1987)

- vocal activity patterns for all $K$ participants, seen together
- only talkspurt start/end times = text-independence
- formally, at time $t$:
  - vocal activity of participant $k$: $q_t[k] \in \mathbb{V} \equiv \{\square, \blacksquare\} \equiv \{0, 1\}$
  - entire $K$-participant conversation: $q_t \in \mathbb{V}^K$
- we'll use a discretized version
Vocal Interaction (Dabbs & Ruback, 1987)

- vocal activity patterns for all $K$ participants, seen together
- only talkspurt start/end times = text-independence
- formally, at time $t$:
  - vocal activity of participant $k$: $q_t[k] \in \mathbb{V} \equiv \{\Box, \blacksquare\} \equiv \{0, 1\}$
  - entire $K$-participant conversation: $q_t \in \mathbb{V}^K$

we’ll use a discretized version
Vocal Interaction (Dabbs & Ruback, 1987)

- vocal activity patterns for all $K$ participants, seen together
- only talkspurt start/end times = text-independence
- formally, at time $t$:
  - vocal activity of participant $k$: $q_t[k] \in \{\Box, \blacksquare\} \equiv \{0, 1\}$
  - entire $K$-participant conversation: $q_t \in \mathbb{V}^K$
- we’ll use a discretized version
Vocal Interaction (Dabbs & Ruback, 1987)

- vocal activity patterns for all $K$ participants, seen together
- only talkspurt start/end times = text-independence
- formally, at time $t$:
  - vocal activity of participant $k$: $q_t[k] \in \{□, ■\} \equiv \{0, 1\}$
  - entire $K$-participant conversation: $q_t \in \bigvee^K$
- we’ll use a discretized version
Vocal Interaction (Dabbs & Ruback, 1987)

- vocal activity patterns for all $K$ participants, seen together
- only talkspurt start/end times = text-independence
- formally, at time $t$:
  - vocal activity of participant $k$: $q_t[k] \in \{\square, \blacksquare\} \equiv \{0, 1\}$
  - entire $K$-participant conversation: $q_t \in \bigvee^K$
- we’ll use a discretized version
Vocal Interaction (Dabbs & Ruback, 1987)

- Vocal activity patterns for all $K$ participants, seen together
- Only talkspurt start/end times = text-independence
- Formally, at time $t$:
  - Vocal activity of participant $k$: $q_t[k] \in \{\square, ■\} \equiv \{0, 1\}$
  - Entire $K$-participant conversation: $q_t \in \mathbb{V}^K$
- We’ll use a discretized version
Vocal Interaction (Dabbs & Ruback, 1987)

- vocal activity patterns for all $K$ participants, seen together
- only talkspurt start/end times = text-independence
- formally, at time $t$:
  - vocal activity of participant $k$: $q_t[k] \in \{\Box, \blacksquare\} \equiv \{0, 1\}$
  - entire $K$-participant conversation: $q_t \in \mathbb{V}^K$

we’ll use a discretized version
Vocal Interaction (Dabbs & Ruback, 1987)

- vocal activity patterns for all $K$ participants, seen together
- only talkspurt start/end times = text-independence
- formally, at time $t$:
  - vocal activity of participant $k$: $q_t[k] \in \{\square, \blacksquare\} \equiv \{0, 1\}$
  - entire $K$-participant conversation: $q_t \in \nabla^K$
- we’ll use a discretized version
Vocal Interaction (Dabbs & Ruback, 1987)

- vocal activity patterns for all $K$ participants, **seen together**
- only talkspurt start/end times = text-independence
- formally, at time $t$:
  - vocal activity of participant $k$: $q_t[k] \in \mathbb{V} \equiv \{\Box, \blacksquare\} \equiv \{0, 1\}$
  - entire $K$-participant conversation: $q_t \in \mathbb{V}^K$
- we’ll use a discretized version
Variability: Several Questions

1. Do participants to multi-party conversation vary in their exhibited preferences of relative talkspurt deployment timing?
   - Trivial: obviously.
2. Are their preferences predicted by their identity?
3. Are their preferences predictive of their identity?

K. Laskowski & T. Schultz
Interspeech 2008, Brisbane, Australia
Variability: Several Questions

1. Do participants to multi-party conversation vary in their exhibited preferences of relative talkspurt deployment timing?
   - Trivial: obviously.

2. Are their preferences predicted by their identity?

3. Are their preferences predictive of their identity?
Variability: Several Questions

1. Do participants to multi-party conversation vary in their exhibited preferences of relative talkspurt deployment timing?
   - **Trivial**: obviously.

2. Are their preferences *predicted* by their identity?
3. Are their preferences *predictive* of their identity?
Variability: Several Questions

1. Do participants in multi-party conversation vary in their exhibited preferences of relative talkspurt deployment timing?
   - Trivial: obviously.

2. Are their preferences predicted by their identity?

3. Are their preferences predictive of their identity?
Variability: Several Questions

1. Do participants to multi-party conversation vary in their exhibited preferences of relative talkspurt deployment timing?
   - Trivial: obviously.

2. Are their preferences *predicted by* their identity?

3. Are their preferences *predictive of* their identity?
Why Do This?

- contrast **participant class** profiles with **participant** profiles
  - social psychology predicts that preferences of relative timing in talkspurt deployment are predictive of speaker’s place in social hierarchy
  - recent progress, computationally
  - not known to what extent classifiers are detecting specific participants

- potential case for participant adaptation at low-level, early processing stages of conversation understanding systems, ie. vocal activity detection
Why Do This?

1. contrast **participant class** profiles with **participant** profiles

- social psychology predicts that preferences of relative timing in talkspurt deployment are predictive of speaker’s place in social hierarchy
- recent progress, computationally
- not known to what extent classifiers are detecting specific participants

2. potential case for participant adaptation at low-level, early processing stages of conversation understanding systems, ie. vocal activity detection
Why Do This?

1. contrast **participant class** profiles with **participant** profiles
   - social psychology predicts that preferences of relative timing in talkspurt deployment are predictive of speaker's place in social hierarchy
   - recent progress, computationally
   - not known to what extent classifiers are detecting specific participants

2. potential case for participant adaptation at low-level, early processing stages of conversation understanding systems, ie. vocal activity detection
Why Do This?

1. contrast **participant class** profiles with **participant** profiles
   - social psychology predicts that preferences of relative timing in talkspurt deployment are predictive of speaker’s place in social hierarchy
   - recent progress, computationally
     - not known to what extent classifiers are detecting specific participants

2. potential case for participant adaptation at low-level, early processing stages of conversation understanding systems, ie. vocal activity detection
Why Do This?

1. contrast participant class profiles with participant profiles
   - social psychology predicts that preferences of relative timing in talkspurt deployment are predictive of speaker’s place in social hierarchy
   - recent progress, computationally
   - not known to what extent classifiers are detecting specific participants

2. potential case for participant adaptation at low-level, early processing stages of conversation understanding systems, ie. vocal activity detection
Formulation of the Problem

- Problem 1: participant identities known (but not assigned)
  - attribute each of $K$ identities to one of $K$ channels
- Problem 2: participant identities not known
Formulation of the Problem

- **Problem 1**: participant identities known (but not assigned)
  - Attribute each of $K$ identities to one of $K$ channels

- **Problem 2**: participant identities not known
  - Draw $K$ identities from a population of $N \gg K$
  - Attribute each of $K$ drawn identities to one of $K$ channels
Formulation of the Problem

- **Problem 1:** participant identities known (but not assigned)
  - attribute each of $K$ identities to one of $K$ channels

- **Problem 2:** participant identities not known
  - draw $K$ identities from a population of $N \gg K$
  - attribute each of $K$ drawn identities to one of $K$ channels
Formulation of the Problem

- **Problem 1:** participant identities known (but not assigned)
  1. attribute each of $K$ identities to one of $K$ channels

- **Problem 2:** participant identities not known
  1. draw $K$ identities from a population of $N \gg K$
  2. attribute each of $K$ drawn identities to one of $K$ channels
Formulation of the Problem

- Problem 1: participant identities known (but not assigned)
  1. attribute each of $K$ identities to one of $K$ channels

- Problem 2: participant identities not known
  1. draw $K$ identities from a population of $N \gg K$
  2. attribute each of $K$ drawn identities to one of $K$ channels
Formulation of the Problem

- Problem 1: participant identities known (but not assigned)
  1. attribute each of $K$ identities to one of $K$ channels

- Problem 2: participant identities not known
  1. draw $K$ identities from a population of $N \gg K$
  2. attribute each of $K$ drawn identities to one of $K$ channels
Related Computational Work on Meetings

Static characterization using long-term (entire meeting) observation of vocal interaction:

- of meeting participants
  - dominance rankings: Rienks & Heylen, MLMI 2005
  - influence rankings: Rienks et al., ICMI 2006
  - seniority: Laskowski et al., SIGdial 2008
  - roles: Favre et al., ICMI 2008 (to appear)

- of meetings
  - meeting types: Laskowski et al., SIGdial 2007
Related Computational Work on Meetings

Static characterization using long-term (entire meeting) observation of vocal interaction:

1. of meeting participants
   - dominance rankings: Rienks & Heylen, MLMI 2005
   - influence rankings: Rienks et al., ICMI 2006
   - seniority: Laskowski et al., SIGdial 2008
   - roles: Favre et al., ICMI 2008 (to appear)

2. of meetings
   - meeting types: Laskowski et al., SIGdial 2007
Related Computational Work on Meetings

Static characterization using long-term (entire meeting) observation of vocal interaction:

- of meeting participants
  - dominance rankings: Rienks & Heylen, MLMI 2005
  - influence rankings: Rienks et al., ICMI 2006
  - seniority: Laskowski et al., SIGdial 2008
  - roles: Favre et al., ICMI 2008 (to appear)

- of meetings
  - meeting types: Laskowski et al., SIGdial 2007
Related Computational Work on Meetings

Static characterization using long-term (entire meeting) observation of vocal interaction:

1. of meeting participants
   - dominance rankings: Rienks & Heylen, MLMI 2005
   - influence rankings: Rienks et al., ICMI 2006
   - seniority: Laskowski et al., SIGdial 2008
   - roles: Favre et al., ICMI 2008 (to appear)

2. of meetings
   - meeting types: Laskowski et al., SIGdial 2007
Related Computational Work on Meetings

Static characterization using long-term (entire meeting) observation of vocal interaction:

1. of meeting participants
   - dominance rankings: Rienks & Heylen, MLMI 2005
   - influence rankings: Rienks et al., ICMI 2006
   - seniority: Laskowski et al., SIGdial 2008
   - roles: Favre et al., ICMI 2008 (to appear)

2. of meetings
   - meeting types: Laskowski et al., SIGdial 2007
Related Computational Work on Meetings

Static characterization using long-term (entire meeting) observation of vocal interaction:

1. of meeting participants
   - dominance rankings: Rienks & Heylen, MLMI 2005
   - influence rankings: Rienks et al., ICMI 2006
   - seniority: Laskowski et al., SIGdial 2008
   - roles: Favre et al., ICMI 2008 (to appear)

2. of meetings
   - meeting types: Laskowski et al., SIGdial 2007
Related Computational Work on Meetings

Static characterization using long-term (entire meeting) observation of vocal interaction:

1. of meeting participants
   - dominance rankings: Rienks & Heylen, MLMI 2005
   - influence rankings: Rienks et al., ICMI 2006
   - seniority: Laskowski et al., SIGdial 2008
   - roles: Favre et al., ICMI 2008 (to appear)

2. of meetings
   - meeting types: Laskowski et al., SIGdial 2007
Classifying Participants Independently

1. cannot model interaction with *specific* other participants
2. feature space with non-specific others may be non-convex
3. may require recombination heuristics
4. a participant may be assigned to $\geq 2$ channels

K. Laskowski & T. Schultz
Interspeech 2008, Brisbane, Australia
Classifying Participants Independently

1. cannot model interaction with specific other participants
2. feature space with non-specific others may be non-convex
3. may require recombination heuristics
4. a participant may be assigned to $\geq 2$ channels

K. Laskowski & T. Schultz
Interspeech 2008, Brisbane, Australia
Classifying Participants Independently

- Cannot model interaction with specific other participants
- Feature space with non-specific others may be non-convex
- May require recombination heuristics
- A participant may be assigned to $\geq 2$ channels
Classifying Participants Independently

- cannot model interaction with specific other participants
- feature space with non-specific others may be non-convex
- may require recombination heuristics
- a participant may be assigned to ≥2 channels
Classifying Participants Independently

1. cannot model interaction with specific other participants
2. feature space with non-specific others may be non-convex
3. may require recombination heuristics
4. a participant may be assigned to $\geq 2$ channels

Bob
Des
Ann
Joe
Classifying Participants Independently

Problems:

1. cannot model interaction with specific other participants
2. feature space with non-specific others may be non-convex
3. may require recombination heuristics
4. a participant may be assigned to ≥2 channels
Classifying Participants Independently

Problems:
1. cannot model interaction with specific other participants
   - feature space with non-specific others may be non-convex
2. may require recombination heuristics
   - a participant may be assigned to ≥2 channels
Classifying Participants Independently

Problems:

1. cannot model interaction with **specific** other participants
   - feature space with non-specific others may be non-convex
2. may require recombination heuristics
   - a participant may be assigned to $\geq 2$ channels
Classifying Participants Independently

Problems:
1. cannot model interaction with specific other participants
   - feature space with non-specific others may be non-convex
2. may require recombination heuristics
   - a participant may be assigned to \( \geq 2 \) channels

Solution: model participants \textit{jointly}
Recognizing Participants Jointly

F describes interaction between all K participants
Recognizing Participants Jointly

F describes interaction between all $K$ participants
Feature Types in F

1. probability of vocalizing (V)
2. probability of initiating vocalization (VI) in prior silence
3. probability of continuing vocalization (VC) in prior non-overlap
4. probability of initiating overlap (OI) in prior non-overlap
5. probability of continuing overlap (OC) in prior overlap
Feature Types in $F$

1. probability of vocalizing ($V$)
2. probability of initiating vocalization ($VI$) in prior silence
3. probability of continuing vocalization ($VC$) in prior non-overlap
4. probability of initiating overlap ($OI$) in prior non-overlap
5. probability of continuing overlap ($OC$) in prior overlap

\[ f_k^V \]
Feature Types in $F$

1. probability of vocalizing (V)
2. probability of initiating vocalization (VI) in prior silence
3. probability of continuing vocalization (VC) in prior non-overlap
4. probability of initiating overlap (OI) in prior non-overlap
5. probability of continuing overlap (OC) in prior overlap

$k$

$f_k^{VI}$
Feature Types in $F$

1. probability of vocalizing (V)
2. probability of initiating vocalization (VI) in prior silence
3. probability of continuing vocalization (VC) in prior non-overlap
4. probability of initiating overlap (OI) in prior non-overlap
5. probability of continuing overlap (OC) in prior overlap

$$f_k^{VC}$$
Feature Types in $F$

1. probability of vocalizing ($V$)
2. probability of initiating vocalization ($VI$) in prior silence
3. probability of continuing vocalization ($VC$) in prior non-overlap
4. probability of initiating overlap ($OI$) in prior non-overlap
5. probability of continuing overlap ($OC$) in prior overlap

\[ f_{k,j}^{OI} \]
Feature Types in $F$

1. probability of vocalizing ($V$)
2. probability of initiating vocalization ($VI$) in prior silence
3. probability of continuing vocalization ($VC$) in prior non-overlap
4. probability of initiating overlap ($OI$) in prior non-overlap
5. probability of continuing overlap ($OC$) in prior overlap

\[ f_{k,j}^{OC} \]
Problem 1: Assigning Known Participant Identities

- know the identities of the $K$ participants,
  \[ G = \{ \text{Ann}, \text{Bob}, \text{Cyp} \} \]
- but don't know which channel each participant is on
  \[ g \in G = \{ [A, B, C], [A, C, B], [B, A, C], \ldots \} \]

**GOAL**: find the correct permutation $g^*$, of $K!$ alternatives

- compute features $F$
- require a model $P(g|F)$ such that

\[
g^* = \arg \max_{g \in G} P(g|F)
\]
Problem 1: Assigning Known Participant Identities

- know the identities of the $K$ participants,
  \[ G = \{ \text{Ann, Bob, Cyp} \} \]
- but don’t know which channel each participant is on
  \[ g \in G = \{ [A, B, C], [A, C, B], [B, A, C], \ldots \} \]

**GOAL**: find the correct permutation $g^*$, of $K!$ alternatives

- compute features $F$
- require a model $P(g | F)$ such that
  \[ g^* = \arg \max_{g \in G} P(g | F) \]
Problem 1: Assigning Known Participant Identities

- know the identities of the $K$ participants,
  \[ G = \{ \text{Ann, Bob, Cyp} \} \]
- but don’t know which channel each participant is on
  \[ g \in G = \{ [A, B, C], [A, C, B], [B, A, C], \ldots \} \]

**GOAL**: find the correct permutation $g^*$, of $K!$ alternatives

- compute features $F$
- require a model $P(g|F)$ such that
  \[ g^* = \arg \max_{g \in G} P(g|F) \]
Problem 1: Assigning Known Participant Identities

- know the identities of the $K$ participants,
  \[ G = \{ \text{ANN, BOB, CYP} \} \]
- but don’t know which channel each participant is on
  \[ g \in G = \{ [A, B, C], [A, C, B], [B, A, C], \ldots \} \]

**GOAL:** find the correct permutation $g^*$, of $K!$ alternatives

- compute features $F$
- require a model $P(g | F)$ such that
  \[ g^* = \arg \max_{g \in G} P(g | F) \]
Problem 1: Assigning Known Participant Identities

- know the identities of the $K$ participants,
  \[ G = \{\text{Ann, Bob, Cyp}\} \]
- but don’t know which channel each participant is on
  \[ g \in G = \{[A, B, C], [A, C, B], [B, A, C], \ldots\} \]

**GOAL**: find the correct permutation $g^*$, of $K!$ alternatives

1. compute features $F$
2. require a model $P(g | F)$ such that
   \[ g^* = \arg \max_{g \in G} P(g | F) \]
Problem 1: Assigning Known Participant Identities

- know the identities of the $K$ participants,
  \[ G = \{ \text{Ann, Bob, Cyp} \} \]
- but don’t know which channel each participant is on
  \[ g \in G = \{ [A, B, C], [A, C, B], [B, A, C], \ldots \} \]

**GOAL**: find the correct permutation $g^*$, of $K!$ alternatives

1. compute features $F$
2. require a model $P(g|F)$ such that

\[ g^* = \arg \max_{g \in G} P(g|F) \]
Problem 1: Assigning Known Participant Identities

• know the identities of the $K$ participants,

$$\mathcal{G} = \{ \text{ANN, BOB, CYP} \}$$

• but don’t know which channel each participant is on

$$\mathbf{g} \in \mathcal{G} = \{ [A, B, C], [A, C, B], [B, A, C], \ldots \}$$

**GOAL**: find the correct permutation $\mathbf{g}^*$, of $K!$ alternatives

1. compute features $\mathbf{F}$

2. require a model $P(\mathbf{g} | \mathbf{F})$ such that

$$\mathbf{g}^* = \arg \max_{\mathbf{g} \in \mathcal{G}} P(\mathbf{g} | \mathbf{F})$$
Problem 1: Assigning Known Participant Identities

- know the identities of the $K$ participants,
  \[ G = \{ \text{Ann}, \text{Bob}, \text{Cyp} \} \]
- but don’t know which channel each participant is on
  \[ g \in G = \{ [A, B, C], [A, C, B], [B, A, C], \ldots \} \]

**GOAL**: find the correct permutation $g^*$, of $K!$ alternatives

1. compute features $F$
2. require a model $P(g \mid F)$ such that

\[ g^* = \arg \max_{g \in G} P(g \mid F) \]
Problem 2: Assigning Unknown Participant Identities

- do not know the identities of the $K$ participants,

$$G \in \mathcal{P} = \{\text{Ann}, \text{Bob}, \text{Cyp}, \text{Des}, \text{Edi}, \cdots\}$$

- must draw $K$ from $||\mathcal{P}|| \gg ||G||$ alternatives

**GOAL**: find the correct set $G$ and its correct permutation $g^*$

- compute features $F$

- require a model $P(g|F)$ such that

$$g^* = \arg \max_{g \in G} P(g|F)$$

- but now the arg max may be intractable
Problem 2: Assigning Unknown Participant Identities

- do not know the identities of the $K$ participants,
  \[ G \in \mathcal{P} = \{ \text{Ann, Bob, Cyp, Des, Edi, } \ldots \} \]
- must draw $K$ from $||\mathcal{P}|| \gg ||G||$ alternatives

**GOAL**: find the correct set $G$ and its correct permutation $g^*$

1. compute features $F$
2. require a model $P(g|F)$ such that

\[ g^* = \arg \max_{g \in G} P(g|F) \]

3. but now the arg max may be intractable
Problem 2: Assigning Unknown Participant Identities

- do not know the identities of the $K$ participants,
  $$G \in \mathcal{P} = \{\text{Ann, Bob, Cyp, Des, Edi, \ldots}\}$$
- must draw $K$ from $||\mathcal{P}|| \gg ||G||$ alternatives

**GOAL**: find the correct set $G$ and its correct permutation $g^*$

1. compute features $F$
2. require a model $P(g|F)$ such that

$$g^* = \arg \max_{g \in G} P(g|F)$$

but now the arg max may be intractable
Problem 2: Assigning Unknown Participant Identities

- do not know the identities of the $K$ participants,
  \[ G \in \mathcal{P} = \{ \text{Ann, Bob, Cyp, Des, Edi,} \ldots \} \]
- must draw $K$ from $\|\mathcal{P}\| \gg \|G\|$ alternatives

**GOAL**: find the correct set $G$ and its correct permutation $g^*$

1. compute features $F$
2. require a model $P(g|F)$ such that
   \[ g^* = \arg\max_{g \in \mathcal{G}} P(g|F) \]
3. but now the arg max may be intractable
Problem 2: Assigning Unknown Participant Identities

- do not know the identities of the $K$ participants,
  \[ G \in \mathcal{P} = \{\text{Ann, Bob, Cyp, Des, Edi, } \ldots \} \]
- must draw $K$ from $\|\mathcal{P}\| \gg \|G\|$ alternatives

**GOAL**: find the correct set $G$ and its correct permutation $g^*$

1. compute features $F$
2. require a model $P(g|F)$ such that

\[
g^* = \arg \max_{g \in G} P(g|F)
\]

but now the arg max may be intractable
Problem 2: Assigning Unknown Participant Identities

- do not know the identities of the $K$ participants,
  $$\mathcal{G} \in \mathcal{P} = \{\text{Ann, Bob, Cyp, Des, Edi, \ldots}\}$$
- must draw $K$ from $|\mathcal{P}| \gg |\mathcal{G}|$ alternatives

**GOAL**: find the correct set $\mathcal{G}$ and its correct permutation $g^*$

1. compute features $F$
2. require a model $P( g | F )$ such that
   $$g^* = \arg \max_{g \in \mathcal{G}} P( g | F )$$
3. but now the arg max may be intractable
The Need for a Greedy Search

- assuming (1) a finite number $|\mathcal{P}|$ of candidate participants,
- and (2) existence of a non-unique UNK participant,
- the number of candidate $K$-assignments is

$$|\mathcal{G}| = \sum_{j=0}^{K} \frac{K!}{(K-j)!j!} \cdot \frac{(|\mathcal{P}| - 1)!}{(|\mathcal{P}| - 1 - j)!}$$

Proposed Search Algorithm:

1. set $g[k] = \text{UNK}$, for all $1 \leq k \leq K$
2. try each candidate in $\mathcal{P}$, in each UNK position in $g$
3. maximize $P(g | F)$
The Need for a Greedy Search

- assuming (1) a finite number $|\mathcal{P}|$ of candidate participants,
- and (2) existence of a non-unique UNK participant,

the number of candidate $K$-assignments is

$$|G| = \sum_{j=0}^{K} \frac{K!}{(K-j)!j!} \cdot \frac{(|\mathcal{P}| - 1)!}{(|\mathcal{P}| - 1 - j)!}$$

Proposed Search Algorithm:

1. set $g[k] = \text{UNK}$, for all $1 \leq k \leq K$
2. try each candidate in $\mathcal{P}$, in each UNK position in $g$
3. maximize $P(g | F)$
The Need for a Greedy Search

- assuming (1) a finite number $|\mathcal{P}|$ of candidate participants,
- and (2) existence of a non-unique UNK participant,
- the number of candidate $K$-assignments is

$$|G| = \sum_{j=0}^{K} \frac{K!}{(K-j)!j!} \cdot \frac{(|\mathcal{P}| - 1)!}{(|\mathcal{P}| - 1 - j)!}$$

Proposed Search Algorithm:

1. set $g[k] = \text{UNK}$, for all $1 \leq k \leq K$
2. try each candidate in $\mathcal{P}$, in each $\text{UNK}$ position in $g$
3. maximize $P(g | F)$
The Need for a Greedy Search

- assuming (1) a finite number $|\mathcal{P}|$ of candidate participants,
- and (2) existence of a non-unique $\text{UNK}$ participant,
- the number of candidate $K$-assignments is

$$|G| = \sum_{j=0}^{K} \frac{K!}{(K-j)!j!} \cdot \frac{(|\mathcal{P}|-1)!}{(|\mathcal{P}|-1-j)!}$$

Proposed Search Algorithm:

1. set $g[k] = \text{UNK}$, for all $1 \leq k \leq K$
2. try each candidate in $\mathcal{P}$, in each $\text{UNK}$ position in $g$
3. maximize $P(g|F)$
The Need for a Greedy Search

- assuming (1) a finite number $|\mathcal{P}|$ of candidate participants,
- and (2) existence of a non-unique $\text{UNK}$ participant,
- the number of candidate $K$-assignments is

$$|\mathcal{G}| = \sum_{j=0}^{K} \frac{K!}{(K-j)!j!} \cdot \frac{(|\mathcal{P}| - 1)!}{(|\mathcal{P}| - 1 - j)!}$$

Proposed Search Algorithm:

1. set $g[k] = \text{UNK}$, for all $1 \leq k \leq K$
2. try each candidate in $\mathcal{P}$, in each $\text{UNK}$ position in $g$
3. maximize $P(g | F)$
The Need for a Greedy Search

- assuming (1) a finite number $|\mathcal{P}|$ of candidate participants,
- and (2) existence of a non-unique $\text{UNK}$ participant,
- the number of candidate $K$-assignments is

$$|G| = \sum_{j=0}^{K} \frac{K!}{(K-j)!j!} \cdot \frac{(|\mathcal{P}| - 1)!}{(|\mathcal{P}| - 1 - j)!}$$

Proposed Search Algorithm:

1. set $g[k] = \text{UNK}$, for all $1 \leq k \leq K$
2. try each candidate in $\mathcal{P}$, in each $\text{UNK}$ position in $g$
3. maximize $P\left( g \mid \mathbf{F} \right)$
The Model $P(\mathbf{g} | \mathbf{F})$

\[
\mathbf{g}^* = \arg \max_{\mathbf{g} \in \mathcal{G}} P(\mathbf{g} | \mathbf{F}) \\
= \arg \max_{\mathbf{g} \in \mathcal{G}} \underbrace{P(\mathbf{g})}_{\text{MM}} \underbrace{P(\mathbf{F} | \mathbf{g})}_{\text{BM}}
\]

\[
P(\mathbf{g}) = \prod_{k=1}^{K} P(\mathbf{g}[k])
\]

\[
P(\mathbf{F} | \mathbf{g}) = \prod_{k=1}^{K} P(f_k | \theta_{\mathbf{g}[k]}) \prod_{j \neq k}^{K} P(f_{kj} | \theta_{\mathbf{g}[k], \mathbf{g}[j]})
\]
The Model $P(g | F)$

$$g^* = \arg \max_{g \in G} P(g | F)$$

$$= \arg \max_{g \in G} \left( P(g) \cdot P(F | g) \right)$$

$$P(g) = \prod_{k=1}^{K} P(g[k])$$

$$P(F | g) = \prod_{k=1}^{K} P(f_k | \theta_{g[k]}) \prod_{j \neq k} P(f_{kj} | \theta_{g[k], g[j]})$$
The Model $P(g | F)$

\[
g^* = \arg \max_{g \in G} P(g | F) = \arg \max_{g \in G} \left( P(g) \times P(F | g) \right)
\]

\[
P(g) = \prod_{k=1}^{K} P(g[k])
\]

\[
P(F | g) = \prod_{k=1}^{K} P(f_k | \theta_{g[k]}) \prod_{j \neq k} P(f_{kj} | \theta_{g[k],g[j]})
\]
Data

- **ICSI Meeting Corpus** (Janin et al, 2003)
  - naturally occurring, $3 \leq K \leq 9$
  - **TrainSet**: 33 meetings
  - **DevSet**: 18 meetings
  - **EvalSet**: 16 meetings
  - 14 participants occur $\geq 7$ times in **TrainSet**,
    $$P = \{ S_1, S_2, \cdots, S_{13}, S_{14}, \text{UNK} \}$$
  - time-aligned segmentations for all meetings
Data

- **ICSI Meeting Corpus** (*Janin et al., 2003*)
  - naturally occurring, $3 \leq K \leq 9$
  - **TrainSet**: 33 meetings
  - **DevSet**: 18 meetings
  - **EvalSet**: 16 meetings
  - 14 participants occur $\geq 7$ times in **TrainSet**,
    \[ P = \{S_1, S_2, \ldots, S_{13}, S_{14}, \text{Unk}\} \]
  - time-aligned segmentations for all meetings
    - talkspurts, from ICSI MRDA Corpus (*Shriberg et al., 2004*)
    - laugh bouts (*Laskowski & Burger, 2007*)

K. Laskowski & T. Schultz  Interspeech 2008, Brisbane, Australia
Data

- ICSI Meeting Corpus (Janin et al., 2003)
  - naturally occurring, $3 \leq K \leq 9$
  - TrainSet: 33 meetings
  - DevSet: 18 meetings
  - EvalSet: 16 meetings
- 14 participants occur $\geq 7$ times in TrainSet,
  \[ \mathcal{P} = \{S_1, S_2, \cdots, S_{13}, S_{14}, \text{UNK}\} \]
- time-aligned segmentations for all meetings
- talkspurts, from ICSI MRDA Corpus (Shriberg et al., 2004)
- laugh bouts (Laskowski & Burger, 2007)
## Data

- **ICSI Meeting Corpus** (Janin *et al.*, 2003)
  - naturally occurring, $3 \leq K \leq 9$
  - **TrainSet**: 33 meetings
  - **DevSet**: 18 meetings
  - **EvalSet**: 16 meetings
- 14 participants occur $\geq 7$ times in **TrainSet**,
  
  \[ P = \{S_1, S_2, \cdots, S_{13}, S_{14}, \text{UNK}\} \]
- time-aligned segmentations for all meetings
  - talkspurts, from ICSI MRDA Corpus (*Shriberg* *et al.*, 2004)
  - laugh bouts (*Laskowski & Burger*, 2007)
Data

- ICSI Meeting Corpus (Janin et al, 2003)
  - naturally occurring, $3 \leq K \leq 9$
  - TrainSet: 33 meetings
  - DevSet: 18 meetings
  - EvalSet: 16 meetings
- 14 participants occur $\geq 7$ times in TrainSet,
  \[ P = \{S_1, S_2, \cdots, S_{13}, S_{14}, \text{UNK}\} \]
- time-aligned segmentations for all meetings
  - talkspurts, from ICSI MRDA Corpus (Shriberg et al, 2004)
  - laugh bouts (Laskowski & Burger, 2007)
Data

- **ICSI Meeting Corpus** (Janin et al, 2003)
  - naturally occurring, $3 \leq K \leq 9$
  - TrainSet: 33 meetings
  - DevSet: 18 meetings
  - EvalSet: 16 meetings
- 14 participants occur $\geq 7$ times in TrainSet,
  \[ P = \{ S_1, S_2, \ldots, S_{13}, S_{14}, \text{UNK} \} \]
- time-aligned segmentations for all meetings
  - talkspurts, from ICSI MRDA Corpus (Shriberg et al, 2004)
  - laugh bouts (Laskowski & Burger, 2007)
DevSet Performance on Problem 1

- always guessing UNK (majority) class: 22.9%
- always guessing non-UNK majority class: 11.9%
- top-5 feature type family combination: 69.5%
DevSet Performance on Problem 1

- always guessing UNK (majority) class: 22.9%
- always guessing non-UNK majority class: 11.9%

- top-5 feature type family combination: 69.5%
DevSet Performance on Problem 1

- always guessing UNK (majority) class: 22.9%
- always guessing non-UNK majority class: 11.9%

<table>
<thead>
<tr>
<th>$\Delta T$ (ms)</th>
<th>speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>55.9</td>
</tr>
</tbody>
</table>

- top-5 feature type family combination: 69.5%
DevSET Performance on Problem 1

- always guessing $\text{UNK}$ (majority) class: 22.9%
- always guessing non-$\text{UNK}$ majority class: 11.9%

<table>
<thead>
<tr>
<th>$\Delta T$ (ms)</th>
<th>speech</th>
<th>laughter</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>55.9</td>
<td>36.4</td>
</tr>
</tbody>
</table>

- top-5 feature type family combination: 69.5%
DevSet Performance on Problem 1

- always guessing UNK (majority) class: 22.9%
- always guessing non-UNK majority class: 11.9%

<table>
<thead>
<tr>
<th>$\Delta T$ (ms)</th>
<th>speech</th>
<th>laughter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>all</td>
<td>no BCs</td>
</tr>
<tr>
<td>50</td>
<td>55.9</td>
<td>57.6</td>
</tr>
</tbody>
</table>

- top-5 feature type family combination: 69.5%

K. Laskowski & T. Schultz  Interspeech 2008, Brisbane, Australia
DevSet Performance on Problem 1

- always guessing UNK (majority) class: 22.9%
- always guessing non-UNK majority class: 11.9%

<table>
<thead>
<tr>
<th>$\Delta T$ (ms)</th>
<th>speech</th>
<th>laughter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>all</td>
<td>no BCs</td>
</tr>
<tr>
<td>50</td>
<td>55.9</td>
<td>57.6</td>
</tr>
<tr>
<td>100</td>
<td><strong>60.2</strong></td>
<td>56.8</td>
</tr>
<tr>
<td>200</td>
<td><strong>60.2</strong></td>
<td>53.4</td>
</tr>
<tr>
<td>400</td>
<td>55.1</td>
<td><strong>58.5</strong></td>
</tr>
<tr>
<td>800</td>
<td>47.5</td>
<td>47.5</td>
</tr>
<tr>
<td>1600</td>
<td>54.2</td>
<td>56.8</td>
</tr>
</tbody>
</table>

- top-5 feature type family combination: 69.5%
DevSet Performance on Problem 1

- always guessing UNK (majority) class: 22.9%
- always guessing non-UNK majority class: 11.9%

<table>
<thead>
<tr>
<th>$\Delta T$ (ms)</th>
<th>speech</th>
<th>laughter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>all</td>
<td>no BCs</td>
</tr>
<tr>
<td>50</td>
<td>55.9</td>
<td>57.6</td>
</tr>
<tr>
<td>100</td>
<td>60.2</td>
<td>56.8</td>
</tr>
<tr>
<td>200</td>
<td>60.2</td>
<td>53.4</td>
</tr>
<tr>
<td>400</td>
<td>55.1</td>
<td>58.5</td>
</tr>
<tr>
<td>800</td>
<td>47.5</td>
<td>47.5</td>
</tr>
<tr>
<td>1600</td>
<td>54.2</td>
<td>56.8</td>
</tr>
</tbody>
</table>

- top-5 feature type family combination: 69.5%
Generalization

<table>
<thead>
<tr>
<th>Training</th>
<th>DevSet</th>
<th>EvalSet</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Problem 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TrainSet</td>
<td>69.5</td>
<td>53.9</td>
</tr>
<tr>
<td>TrainSet &amp; DevSet</td>
<td>—</td>
<td>57.8</td>
</tr>
<tr>
<td><strong>Problem 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TrainSet</td>
<td>29.7</td>
<td>30.4</td>
</tr>
<tr>
<td>TrainSet &amp; DevSet</td>
<td>—</td>
<td>34.3</td>
</tr>
</tbody>
</table>

- evident: system optimized on Problem 1 for DevSet
- Problem 2 (drawing from $\mathcal{P}$ and permutation) much harder
- additional training data helps
### Generalization

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th>DevSet</th>
<th>EvalSet</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Problem 1</strong></td>
<td>TRAINSET</td>
<td>69.5</td>
<td>53.9</td>
</tr>
<tr>
<td></td>
<td>TRAINSET &amp; DEVSET</td>
<td>—</td>
<td>57.8</td>
</tr>
<tr>
<td><strong>Problem 2</strong></td>
<td>TRAINSET</td>
<td>29.7</td>
<td>30.4</td>
</tr>
<tr>
<td></td>
<td>TRAINSET &amp; DEVSET</td>
<td>—</td>
<td>34.3</td>
</tr>
</tbody>
</table>

- evident: system optimized on Problem 1 for **DevSet**
- Problem 2 (drawing from $\mathcal{P}$ and permutation) much harder
- additional training data helps
## Generalization

<table>
<thead>
<tr>
<th>Problem 1</th>
<th>TRAINSet</th>
<th>DEVSet</th>
<th>EVALSet</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TRAINSet</td>
<td>69.5</td>
<td>53.9</td>
</tr>
<tr>
<td></td>
<td>TRAINSet &amp; DEVSet</td>
<td>—</td>
<td>57.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Problem 2</th>
<th>TRAINSet</th>
<th>DEVSet</th>
<th>EVALSet</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TRAINSet</td>
<td>29.7</td>
<td>30.4</td>
</tr>
<tr>
<td></td>
<td>TRAINSet &amp; DEVSet</td>
<td>—</td>
<td>34.3</td>
</tr>
</tbody>
</table>

- evident: system optimized on Problem 1 for DEVSet
- Problem 2 (drawing from $\mathcal{P}$ and permutation) much harder
- additional training data helps
### Generalization

<table>
<thead>
<tr>
<th>Problem</th>
<th>Training</th>
<th>DevSet</th>
<th>EvalSet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem 1</td>
<td>TRAINSET</td>
<td>69.5</td>
<td>53.9</td>
</tr>
<tr>
<td></td>
<td>TRAINSET &amp; DEVSET</td>
<td>—</td>
<td>57.8</td>
</tr>
<tr>
<td>Problem 2</td>
<td>TRAINSET</td>
<td>29.7</td>
<td>30.4</td>
</tr>
<tr>
<td></td>
<td>TRAINSET &amp; DEVSET</td>
<td>—</td>
<td>34.3</td>
</tr>
</tbody>
</table>

- evident: system optimized on Problem 1 for DEVSET
- Problem 2 (drawing from $\mathcal{P}$ and permutation) much harder
- additional training data helps
Conclusions

1. relative talkspurt deployment timing preferences
   - are predicted by participant identity
   - are predictive of participant identity (stronger)

2. Problem 1, unseen data
   - participant identities are known
   - must only be shuffled
   - correct prediction in over half of the cases

3. Problem 2, unseen data
   - participant identities are not known
   - must be drawn from larger set and shuffled
   - correct prediction in over a third of the cases
Conclusions

1. Relative talkspurt deployment timing preferences
   - Are predicted by participant identity
   - Are predictive of participant identity (stronger)

2. Problem 1, unseen data
   - Participant identities are known
   - Must only be shuffled
   - Correct prediction in over half of the cases

3. Problem 2, unseen data
   - Participant identities are not known
   - Must be drawn from larger set and shuffled
   - Correct prediction in over a third of the cases
Conclusions

1. relative talkspurt deployment timing preferences
   - are predicted by participant identity
   - are predictive of participant identity (stronger)

2. Problem 1, unseen data
   - participant identities are known
   - must only be shuffled
   - correct prediction in over half of the cases

3. Problem 2, unseen data
Conclusions

1. relative talkspurt deployment timing preferences
   - are predicted by participant identity
   - are predictive of participant identity (stronger)

2. Problem 1, unseen data
   - participant identities are known
   - must only be shuffled
   - correct prediction in over half of the cases

3. Problem 2, unseen data
Conclusions

1. relative talkspurt deployment timing preferences
   - are predicted by participant identity
   - are predictive of participant identity (stronger)

2. Problem 1, unseen data
   - participant identities are known
   - must only be shuffled
   - correct prediction in over half of the cases

3. Problem 2, unseen data
Conclusions

1. Relative talkspurt deployment timing preferences
   - Are predicted by participant identity
   - Are predictive of participant identity (stronger)

2. Problem 1, unseen data
   - Participant identities are known
   - Must only be shuffled
   - Correct prediction in over half of the cases

3. Problem 2, unseen data
   - Participant identities are not known
   - Must be drawn from larger set and shuffled
   - Correct prediction in over a third of the cases
relative talkspurt deployment timing preferences
- are predicted by participant identity
- are predictive of participant identity (stronger)

Problem 1, unseen data
- participant identities are known
- must only be shuffled
- correct prediction in over half of the cases

Problem 2, unseen data
- participant identities are not known
- must be drawn from larger set and shuffled
- correct prediction in over a third of the cases
Conclusions

1. Relative talkspurt deployment timing preferences
   - are predicted by participant identity
   - are predictive of participant identity (stronger)

2. Problem 1, unseen data
   - participant identities are known
   - must only be shuffled
   - correct prediction in over half of the cases

3. Problem 2, unseen data
   - participant identities are not known
   - must be drawn from larger set and shuffled
   - correct prediction in over a third of the cases
Conclusions

1. relative talkspurt deployment timing preferences
   - are predicted by participant identity
   - are predictive of participant identity (stronger)

2. Problem 1, unseen data
   - participant identities are known
   - must only be shuffled
   - correct prediction in over half of the cases

3. Problem 2, unseen data
   - participant identities are **not** known
   - must be drawn from larger set and shuffled
   - correct prediction in over a third of the cases
Conclusions

1. relative talkspurt deployment timing preferences
   - are predicted by participant identity
   - are predictive of participant identity (stronger)

2. Problem 1, unseen data
   - participant identities are known
   - must only be shuffled
   - correct prediction in over half of the cases

3. Problem 2, unseen data
   - participant identities are not known
   - must be drawn from larger set and shuffled
   - correct prediction in over a third of the cases
Future Work

- Always non-UNK majority: 11.8
- Always UNK: 20.6
- Training on TrainSet: 53.9
- Training on TrainSet & DevSet: 57.8

Expected Performance:
- Improved improved MM
- Improved features

Observed Performance
Future Work

- Observed Performance:
  - always non-UNK majority: 11.8
  - always UNK: 20.6
  - train on TrainSet: 30.4
  - train on TrainSet & DevSet: 34.3
  - improved search: 57.8

- Expected Performance:
  - improved features: 53.9

K. Laskowski & T. Schultz
Interspeech 2008, Brisbane, Australia
Future Work

- **Observed Performance**
  - always non-UNK majority: 11.8
  - always UNK: 20.6
  - train on TrainSet: 30.4
  - train on TrainSet & DevSet: 34.3
  - improved search: 57.8
  - improved MM: 53.9

- **Expected Performance**
  - Improved MM
Future Work

- **Observed Performance**
  - always non-UNK majority: 11.8
  - always UNK: 20.6
  - train on TrainSet: 53.9
  - train on TrainSet & DevSet: 57.8

- **Expected Performance**
  - improved search: 34.3
  - improved MM: 30.4
  - improved features: 53.9

K. Laskowski & T. Schultz  Interspeech 2008, Brisbane, Australia
Future Work

- Observed Performance:
  - always non-Unk majority: 11.8
  - always Unk: 20.6
  - train on TrainSet: 30.4
  - train on TrainSet & DevSet: 34.3

- Expected Performance:
  - improved search: 53.9
  - improved MM: 57.8
  - improved features: feature selection

K. Laskowski & T. Schultz | Interspeech 2008, Brisbane, Australia
Thank you for attending.

Thanks also to:

- Liz Shriberg, for access to the ICSI MRDA Corpus