Modeling Vocal Interaction for Text-Independent Participant Characterization in Multi-Party Conversation

Kornel Laskowski[1,2], Mari Ostendorf[3] & Tanja Schultz[1,2]

[1]Cognitive Systems Labs, Universität Karlsruhe
[2]Language Technologies Institute, Carnegie Mellon University
[3]Dept. Electrical Engineering, University of Washington

June 20, 2008

Vocal Interaction (Dabbs & Ruback, 1987)

- vocal activity patterns for all K participants, **seen together**
- talkspurt start/end times = text-independence
- at time t,
- we'll use a discretized version (frame step = 200 ms)

Vocal Interaction (Dabbs & Ruback, 1987)

- vocal activity patterns for all K participants, **seen together**

- talkspurt start/end times = text-independence
- at time t,
  - vocal activity of participant k: $q_t[k] \in V = \{\square, \blacksquare\} = \{0, 1\}$
  - entire K-participant conversation: $q_t \in V^K$
- we'll use a discretized version (frame step = 200 ms)

Vocal Interaction (Dabbs & Ruback, 1987)

- vocal activity patterns for all K participants, **seen together**

- talkspurt start/end times = text-independence
- at time t,
  - vocal activity of participant k: $q_t[k] \in V = \{\square, \blacksquare\} = \{0, 1\}$
  - entire K-participant conversation: $q_t \in V^K$
- we'll use a discretized version (frame step = 200 ms)

Vocal Interaction (Dabbs & Ruback, 1987)

- vocal activity patterns for all K participants, **seen together**

- talkspurt start/end times = text-independence
- at time t,
  - vocal activity of participant k: $\mathbf{q}_t[k] \in \mathbb{V} \equiv \{\square, \blacksquare\} \equiv \{0, 1\}$
  - entire K-participant conversation: $\mathbf{q}_t \in \mathbb{V}^K$
- we'll use a discretized version (frame step = 200 ms)

Vocal Interaction (Dabbs & Ruback, 1987)

- vocal activity patterns for all K participants, **seen together**

- talkspurt start/end times = text-independence
- at time t,
  - vocal activity of participant k: $\mathbf{q}_t[k] \in \mathbb{V} \equiv \{\square, \blacksquare\} \equiv \{0, 1\}$
  - entire K-participant conversation: $\mathbf{q}_t \in \mathbb{V}^K$
- we'll use a discretized version (frame step = 200 ms)

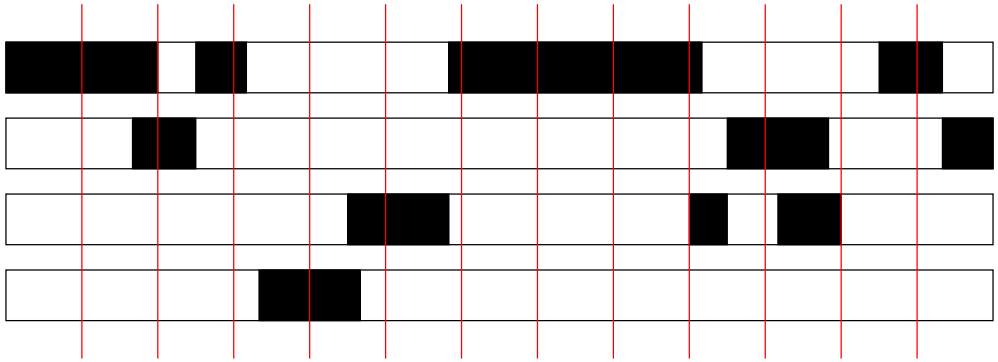
Vocal Interaction (Dabbs & Ruback, 1987)

- vocal activity patterns for all K participants, **seen together**

- talkspurt start/end times = text-independence
- at time t,
  - vocal activity of participant k: $\mathbf{q}_t[k] \in \mathbb{V} \equiv \{\square, \blacksquare\} \equiv \{0, 1\}$
  - entire K-participant conversation: $\mathbf{q}_t \in \mathbb{V}^K$
- we'll use a discretized version (frame step = 200 ms)

Vocal Interaction (Dabbs & Ruback, 1987)

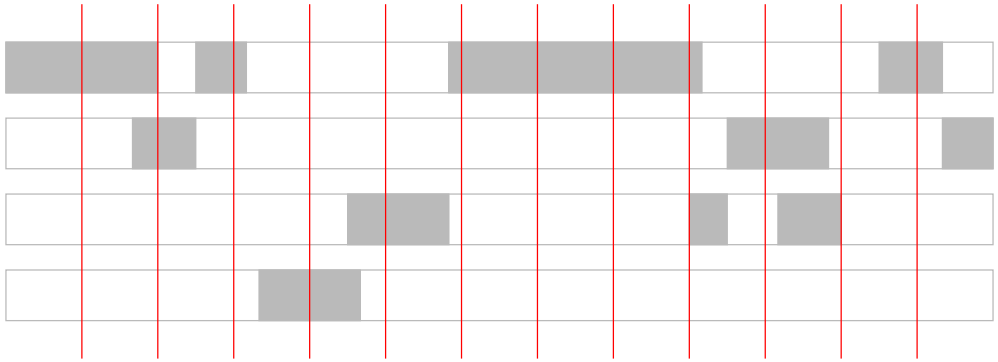
- vocal activity patterns for all K participants, **seen together**



- talkspurt start/end times = text-independence
- at time t,
  - vocal activity of participant k: $\mathbf{q}_t[k] \in \mathbb{V} \equiv \{\square, \blacksquare\} \equiv \{0, 1\}$
  - entire K-participant conversation: $\mathbf{q}_t \in \mathbb{V}^K$
- we'll use a discretized version (frame step = 200 ms)

Vocal Interaction (Dabbs & Ruback, 1987)

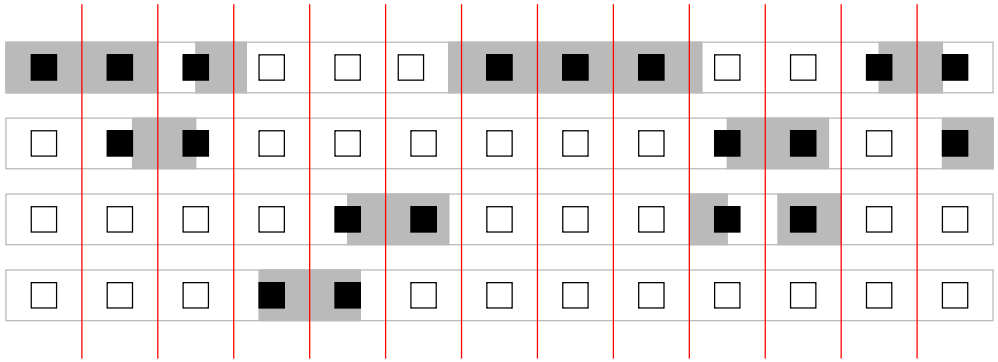
- vocal activity patterns for all K participants, **seen together**



- talkspurt start/end times = text-independence
- at time t,
  - vocal activity of participant k: $\mathbf{q}_t[k] \in \mathbb{V} \equiv \{\square, \blacksquare\} \equiv \{0, 1\}$
  - entire K-participant conversation: $\mathbf{q}_t \in \mathbb{V}^K$
- we'll use a discretized version (frame step = 200 ms)

Vocal Interaction (Dabbs & Ruback, 1987)

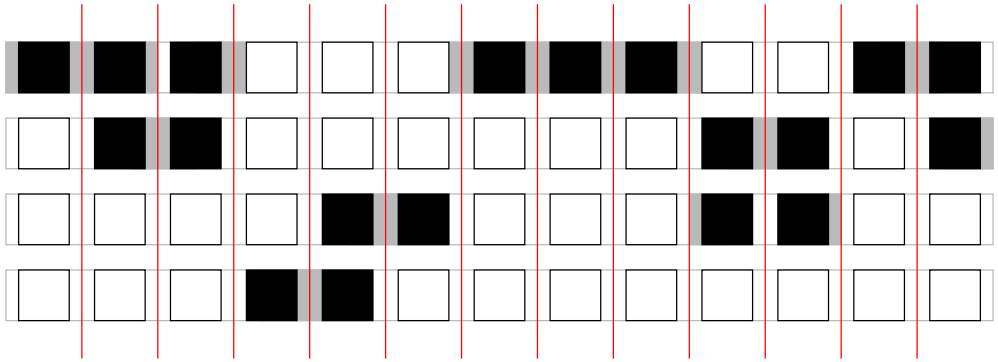
- vocal activity patterns for all K participants, **seen together**



- talkspurt start/end times = text-independence
- at time t,
  - vocal activity of participant k: $\mathbf{q}_t[k] \in \mathbb{V} \equiv \{\square, \blacksquare\} \equiv \{0, 1\}$
  - entire K-participant conversation: $\mathbf{q}_t \in \mathbb{V}^K$
- we'll use a discretized version (frame step = 200 ms)

Vocal Interaction (Dabbs & Ruback, 1987)

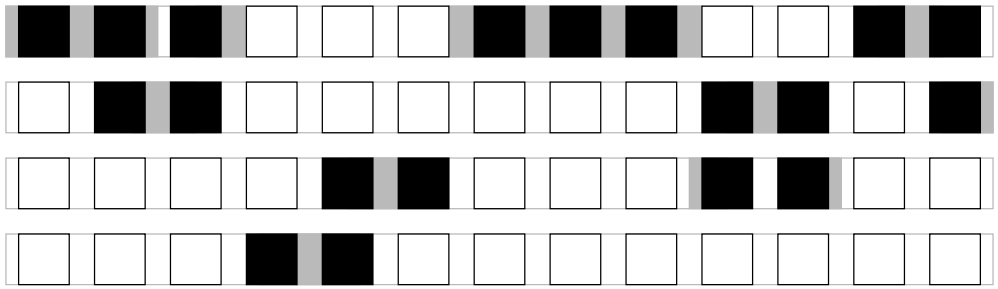
- vocal activity patterns for all K participants, **seen together**



- talkspurt start/end times = text-independence
- at time t,
  - vocal activity of participant k: $\mathbf{q}_t[k] \in \mathbb{V} \equiv \{\square, \blacksquare\} \equiv \{0, 1\}$
  - entire K-participant conversation: $\mathbf{q}_t \in \mathbb{V}^K$
- we'll use a discretized version (frame step = 200 ms)

Vocal Interaction (Dabbs & Ruback, 1987)

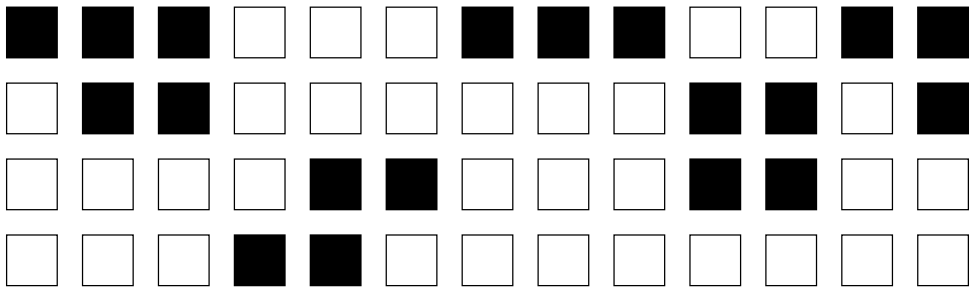
- vocal activity patterns for all K participants, **seen together**



- talkspurt start/end times = text-independence
- at time t,
  - vocal activity of participant k: $\mathbf{q}_t[k] \in \mathbb{V} \equiv \{\square, \blacksquare\} \equiv \{0, 1\}$
  - entire K-participant conversation: $\mathbf{q}_t \in \mathbb{V}^K$
- we'll use a discretized version (frame step = 200 ms)

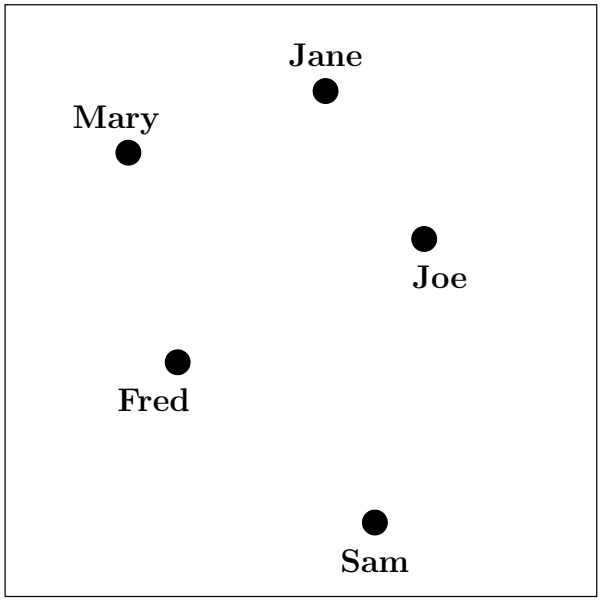
Vocal Interaction (Dabbs & Ruback, 1987)

- vocal activity patterns for all K participants, **seen together**



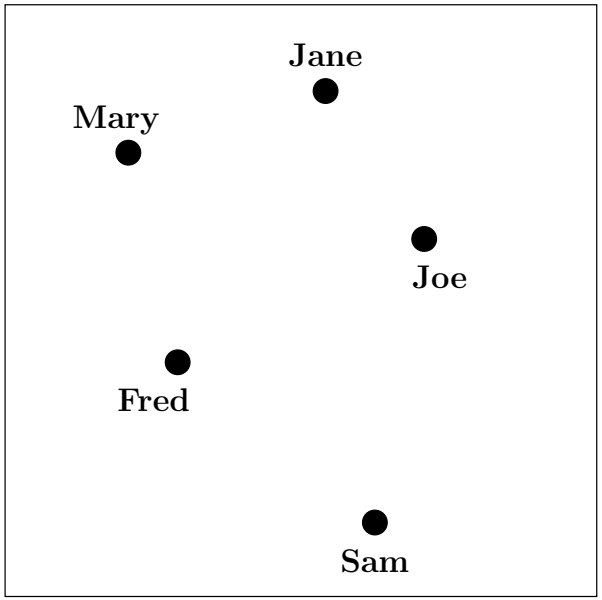
- talkspurt start/end times = text-independence
- at time t,
  - vocal activity of participant k: $\mathbf{q}_t[k] \in \mathbb{V} \equiv \{\square, \blacksquare\} \equiv \{0, 1\}$
  - entire K-participant conversation: $\mathbf{q}_t \in \mathbb{V}^K$
- we'll use a discretized version (frame step = 200 ms)

Participant Characterization



- a *useful* partition of the conversation participants
  - role
  - influence
  - seniority
  - dominance
  - ranking (of the above)
- for all time,

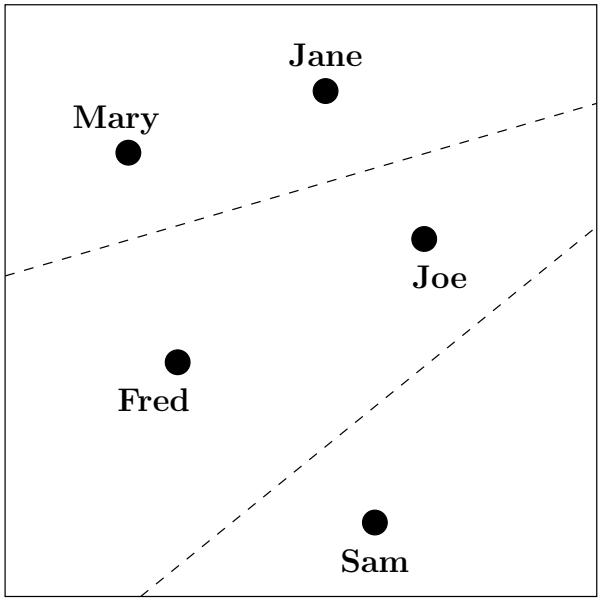
Participant Characterization



- a *useful* partition of the conversation participants
  - role
  - influence
  - seniority
  - dominance
  - ranking (of the above)

- for all time,
  - the class of participant k: $g[k] \in C = \{C_1, \cdots, C_N\}$
  - K participant groups: $g = \{\ \} = h(C)$

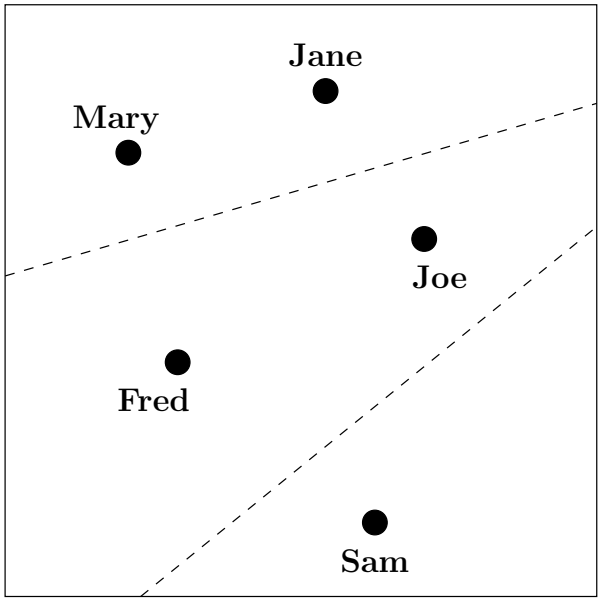
Participant Characterization



- a *useful* partition of the conversation participants
  - role
  - influence
  - seniority
  - dominance
  - ranking (of the above)

- for all time,
  - the class of participant k: $g[k] \in C \equiv \{C_1, \cdots, C_N\}$
  - K-participant group: $\mathbf{g} \in \mathbb{C} \equiv h(C)$

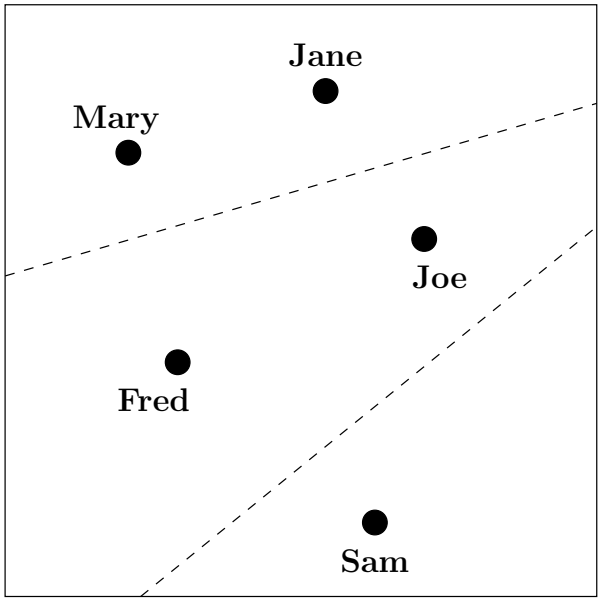
Participant Characterization



- a *useful* partition of the conversation participants
  - role
  - influence
  - seniority
  - dominance
  - ranking (of the above)

- for all time,
  - the class of participant k: $\mathbf{g}[k] \in \mathcal{C} \equiv \{C_1, \cdots, C_N\}$
  - K-participant group: $\mathbf{g} \in \mathbb{C} \equiv h(\mathcal{C})$

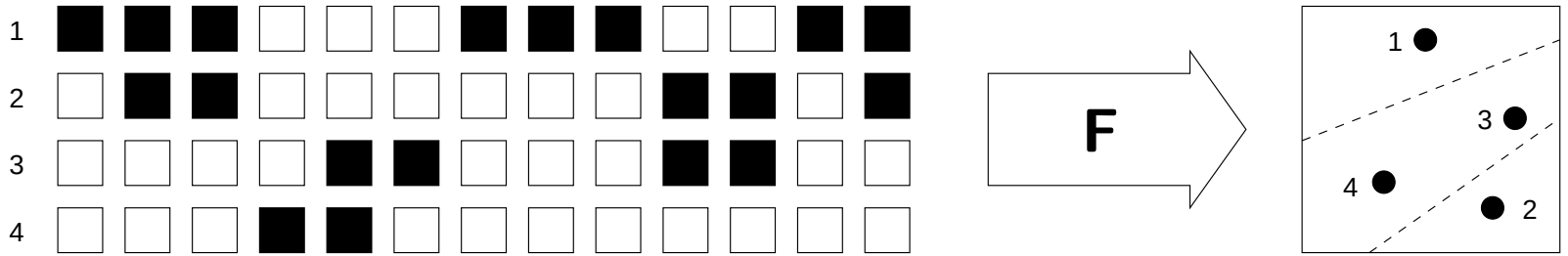
Participant Characterization



- a *useful* partition of the conversation participants
  - role
  - influence
  - seniority
  - dominance
  - ranking (of the above)

- for all time,
  - the class of participant k: $\mathbf{g}[k] \in \mathcal{C} \equiv \{C_1, \cdots, C_N\}$
  - K-participant group: $\mathbf{g} \in \mathbb{C} \equiv h(\mathcal{C})$

What we're trying to do

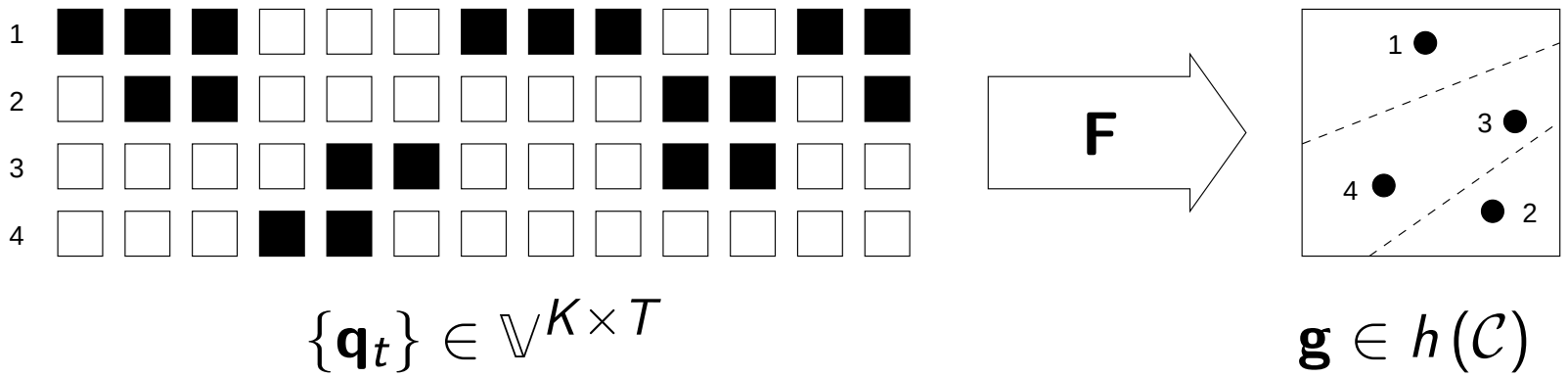


$\{\mathbf{q}_t\} \in \mathbb{V}^{K \times T}$

$\mathbf{g} \in h(\mathcal{C})$

1. given a **sequence** of T K-participant states $\mathbf{q}_t$
2. compute & model features $\mathbf{F}$
3. infer required equivalence classes $\mathbf{g}[k]$ of each participant

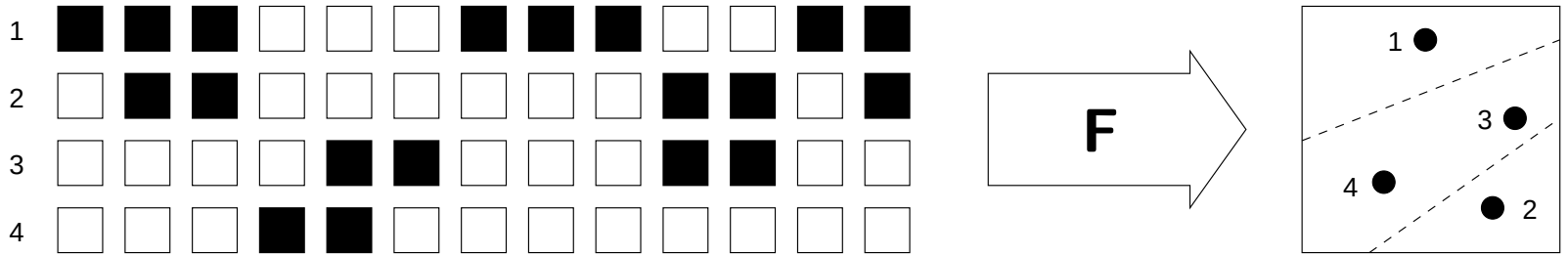
What we're trying to do



$\{\mathbf{q}_t\} \in \mathbb{V}^{K \times T}$ $\mathbf{g} \in h(\mathcal{C})$

1. given a **sequence** of T K-participant states $\mathbf{q}_t$
2. compute & model features $\mathbf{F}$
3. infer required equivalence classes $\mathbf{g}[k]$ of each participant

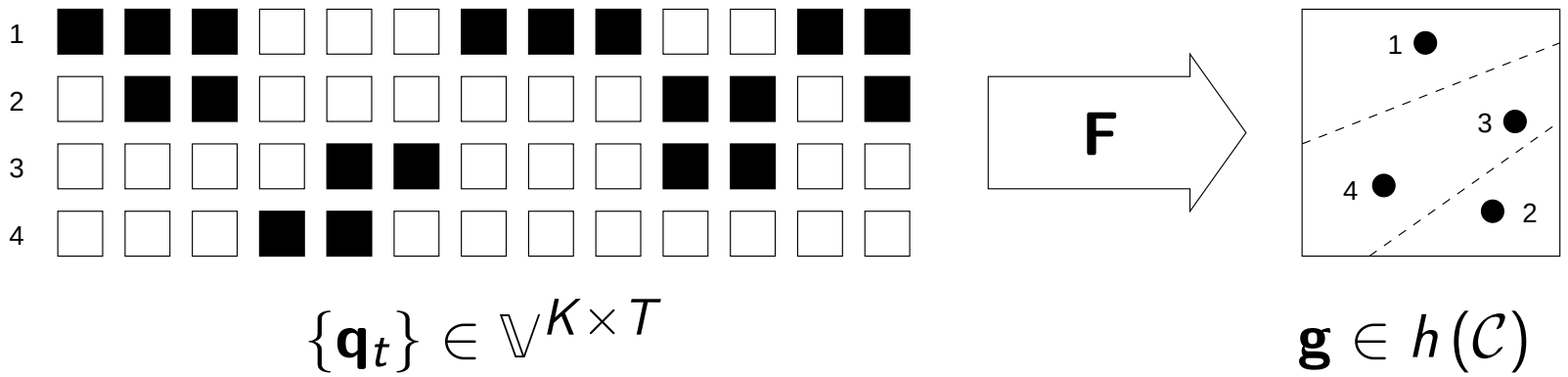
What we're trying to do



$\{\mathbf{q}_t\} \in \mathbb{V}^{K \times T}$ $\quad \mathbf{g} \in h(\mathcal{C})$

1. given a **sequence** of T K-participant states $\mathbf{q}_t$
2. compute & model features **F**
3. infer required equivalence classes $\mathbf{g}[k]$ of each participant

What we're trying to do



$\{\mathbf{q}_t\} \in \mathbb{V}^{K \times T}$

$\mathbf{g} \in h(\mathcal{C})$

1. given a **sequence** of T K-participant states $\mathbf{q}_t$
2. compute & model features **F**
3. infer required equivalence classes $\mathbf{g}[k]$ of each participant

Outline of Talk

0. ... Intro (Motivation & Related Work)
1. Computational Framework
2. Experiments
3. Conclusions

Motivation

- having observed a conversation/meeting, being able to say something about the participants is a **basic competence in conversation understanding**
- lots of research in psycho- and socio-linguistics, 1950-
  - conversation analysis
  - small group research
  - nonverbal interaction
- findings suggest that a participant's talkspurt deployment timing is conditioned on

Motivation

- having observed a conversation/meeting, being able to say something about the participants is a **basic competence in conversation understanding**
- lots of research in psycho- and socio-linguistics, 1950-
  - conversation analysis
  - small group research
  - nonverbal interaction
- findings suggest that a participant's talkspurt deployment timing is conditioned on

Motivation

- having observed a conversation/meeting, being able to say something about the participants is a **basic competence in conversation understanding**
- lots of research in psycho- and socio-linguistics, 1950-
  - converation analysis
  - small group research
  - non-verbal interaction
- findings suggest that a participant's talkspurt deployment timing is conditioned on

Motivation

- having observed a conversation/meeting, being able to say something about the participants is a **basic competence in conversation understanding**
- lots of research in psycho- and socio-linguistics, 1950-
  - converation analysis
  - small group research
  - non-verbal interaction
- findings suggest that a participant's talkspurt deployment timing is conditioned on

Motivation

- having observed a conversation/meeting, being able to say something about the participants is a **basic competence in conversation understanding**
- lots of research in psycho- and socio-linguistics, 1950-
  - converation analysis
  - small group research
  - non-verbal interaction
- findings suggest that a participant's talkspurt deployment timing is conditioned on
  - instrumental status characteristics, e.g. task-specific skills
  - "diffuse status characteristics" e.g. gender, race, etc.

Motivation

- having observed a conversation/meeting, being able to say something about the participants is a **basic competence in conversation understanding**
- lots of research in psycho- and socio-linguistics, 1950-
  - converation analysis
  - small group research
  - non-verbal interaction
- findings suggest that a participant's talkspurt deployment timing is conditioned on
  - instrumental status characteristics, e.g. task-specific skills
  - "diffuse status characteristics", e.g. gender, age, race, etc.

Motivation

- having observed a conversation/meeting, being able to say something about the participants is a **basic competence in conversation understanding**
- lots of research in psycho- and socio-linguistics, 1950-
  - converation analysis
  - small group research
  - non-verbal interaction
- findings suggest that a participant's talkspurt deployment timing is conditioned on
  - instrumental status characteristics, e.g. task-specific skills
  - "diffuse status characteristics", e.g. gender, age, race, etc.

Motivation

- having observed a conversation/meeting, being able to say something about the participants is a **basic competence in conversation understanding**
- lots of research in psycho- and socio-linguistics, 1950-
  - converation analysis
  - small group research
  - non-verbal interaction
- findings suggest that a participant's talkspurt deployment timing is conditioned on
  - instrumental status characteristics, e.g. task-specific skills
  - "diffuse status characteristics", e.g. gender, age, race, etc.

Related Computational Work

- static characterization of meeting participants
  - dominance rankings: Rienks & Heylen, 2005
  - influence rankings: Rienks *et al.*, 2006
- static characterization of radio talk show participants
  - roles: Vinciarelli, 2007
- dynamic characterization of meeting participants
  - roles: Banerjee & Rudnicky, 2004
  - roles: Zancanaro *et al.*, 2006
  - roles: Rienks *et al.*, 2006
- static characterization of conversations
  - meeting types: Laskowski *et al.*, 2007

Related Computational Work

- static characterization of meeting participants
  - dominance rankings: Rienks & Heylen, 2005
  - influence rankings: Rienks *et al.*, 2006
- static characterization of radio talk show participants
  - roles: Vinciarelli, 2007
- dynamic characterization of meeting participants
  - roles: Banerjee & Rudnicky, 2004
  - roles: Zancanaro *et al.*, 2006
  - roles: Rienks *et al.*, 2006
- static characterization of conversations
  - meeting types: Laskowski *et al.*, 2007

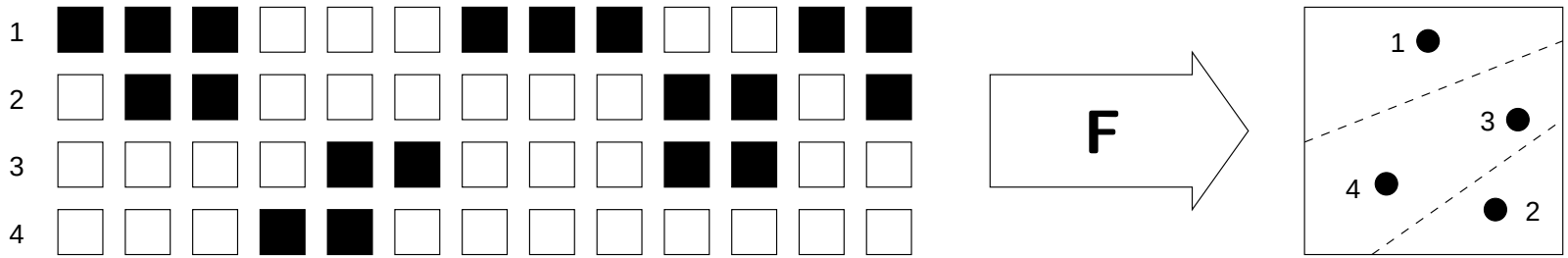
Related Computational Work

- static characterization of meeting participants
  - dominance rankings: Rienks & Heylen, 2005
  - influence rankings: Rienks *et al.*, 2006
- static characterization of radio talk show participants
  - roles: Vinciarelli, 2007
- dynamic characterization of meeting participants
  - roles: Banerjee & Rudnicky, 2004
  - roles: Zancanaro *et al.*, 2006
  - roles: Rienks *et al.*, 2006
- static characterization of conversations
  - meeting types: Laskowski *et al.*, 2007

Related Computational Work

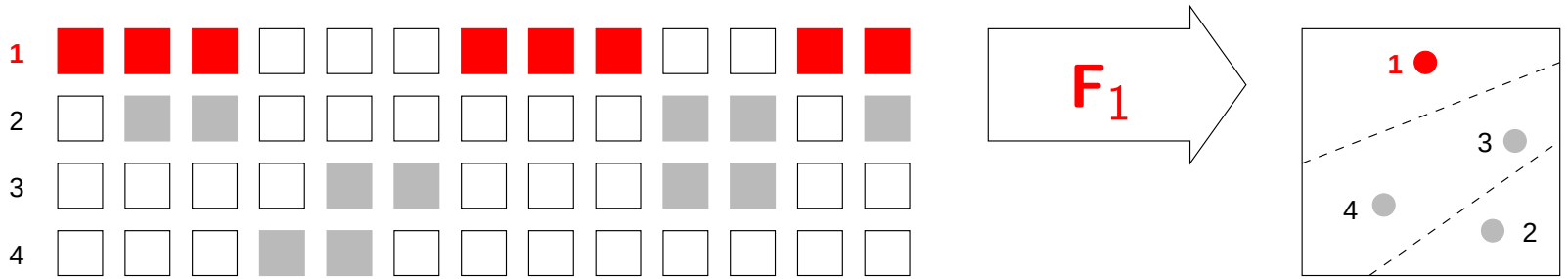
- static characterization of meeting participants
  - dominance rankings: Rienks & Heylen, 2005
  - influence rankings: Rienks *et al.*, 2006
- static characterization of radio talk show participants
  - roles: Vinciarelli, 2007
- dynamic characterization of meeting participants
  - roles: Banerjee & Rudnicky, 2004
  - roles: Zancanaro *et al.*, 2006
  - roles: Rienks *et al.*, 2006
- static characterization of conversations
  - meeting types: Laskowski *et al.*, 2007

Detecting Participant Types Independently



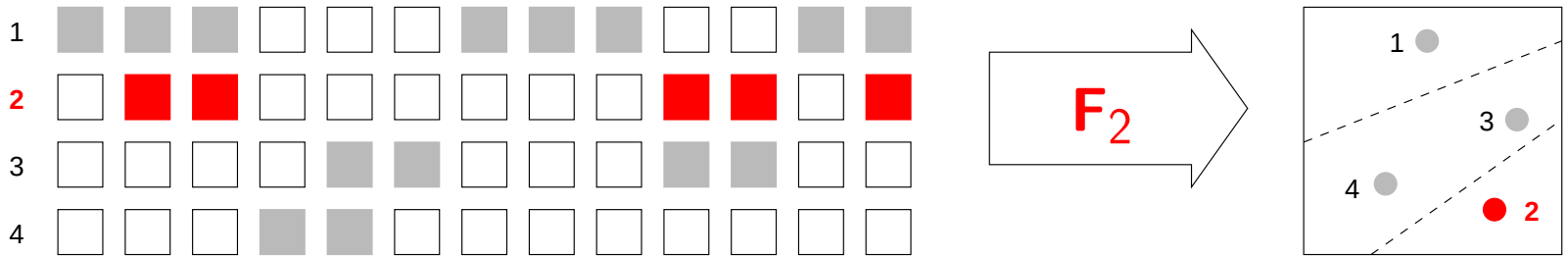
1. cannot model interaction with **specific** other types
   - feature space with non-specific others may be non-convex
2. may require recombination heuristics
   - ≥ 2 participants may be assigned a unique type

Detecting Participant Types Independently



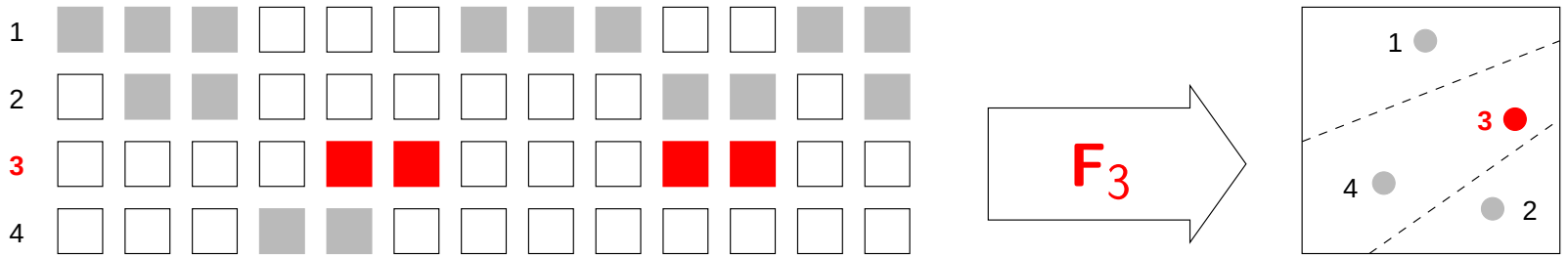
1. cannot model interaction with **specific** other types
   - feature space with non-specific others may be non-convex
2. may require recombination heuristics
   - ≥ 2 participants may be assigned a unique type

Detecting Participant Types Independently



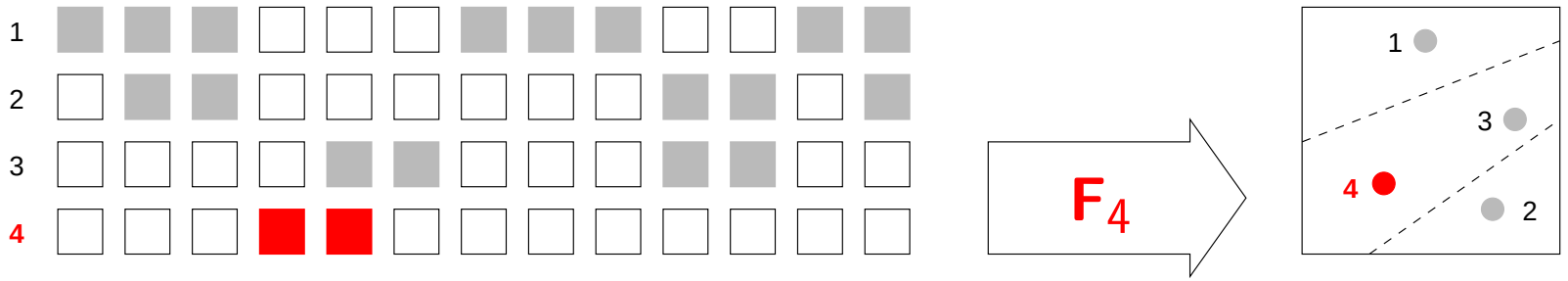
1. cannot model interaction with **specific** other types
   - feature space with non-specific others may be non-convex
2. may require recombination heuristics
   - ≥ 2 participants may be assigned a unique type

Detecting Participant Types Independently



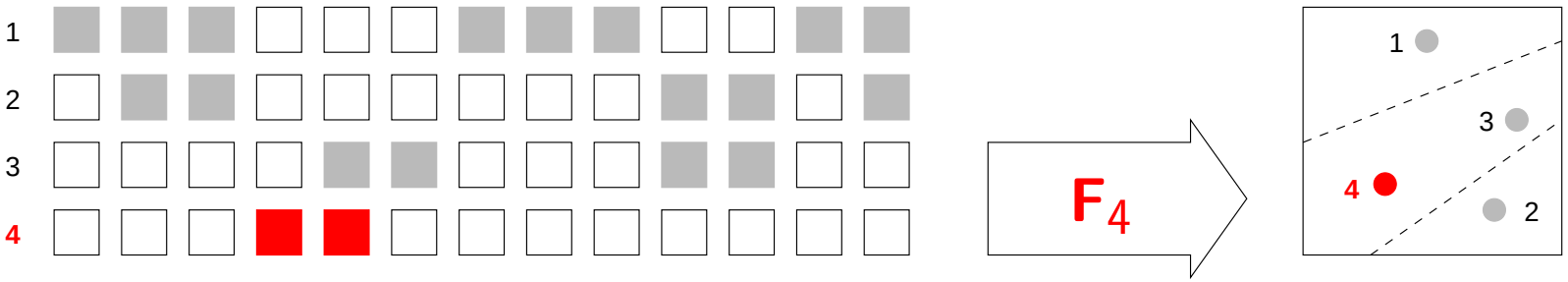
1. cannot model interaction with **specific** other types
   - feature space with non-specific others may be non-convex
2. may require recombination heuristics
   - ≥ 2 participants may be assigned a unique type

Detecting Participant Types Independently



1. cannot model interaction with **specific** other types
   - feature space with non-specific others may be non-convex
2. may require recombination heuristics
   - ≥ 2 participants may be assigned a unique type

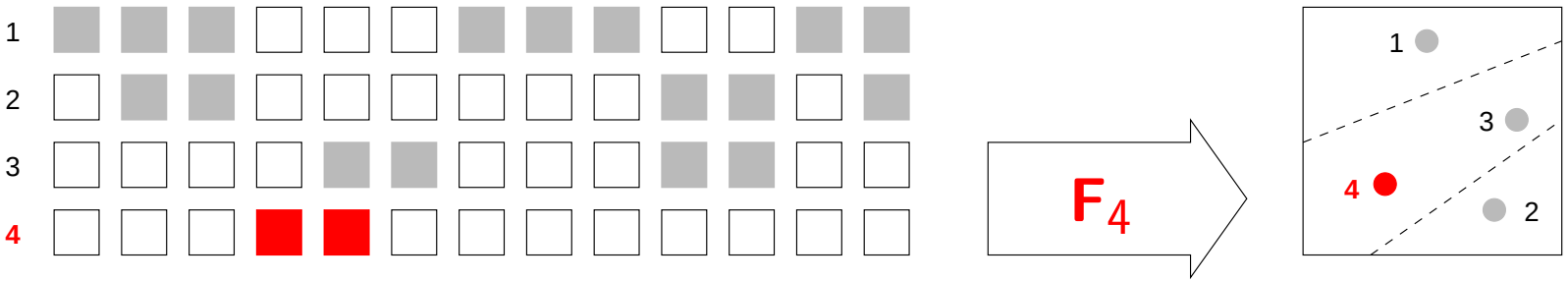
Detecting Participant Types Independently



Problems:

1. cannot model interaction with **specific** other types
   - feature space with non-specific others may be non-convex
2. may require recombination heuristics
   - $\geq$2 participants may be assigned a unique type

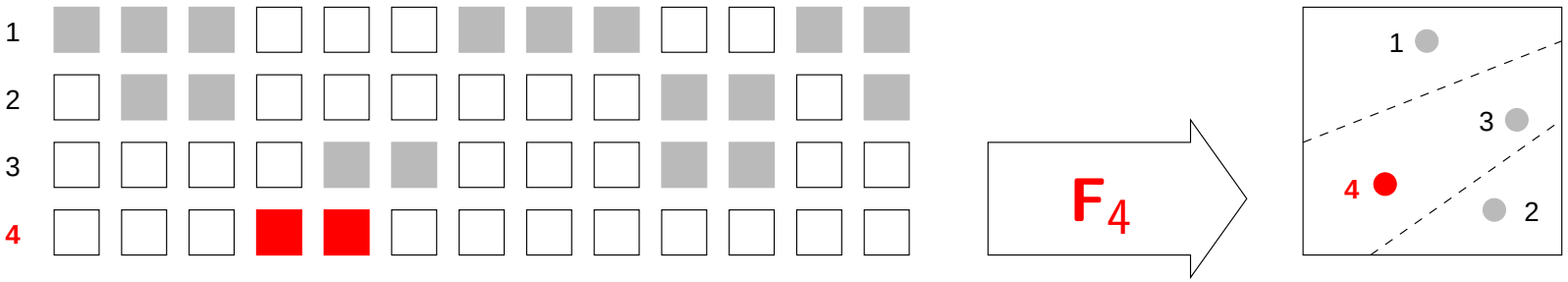
Detecting Participant Types Independently



Problems:

1. cannot model interaction with **specific** other types
   - feature space with non-specific others may be non-convex
2. may require recombination heuristics
   - $\geq$2 participants may be assigned a unique type

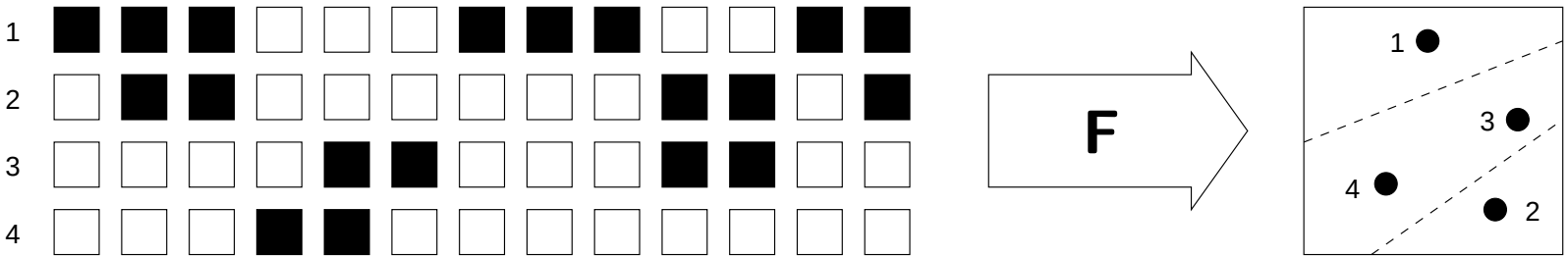
Detecting Participant Types Independently



Problems:

1. cannot model interaction with **specific** other types
   - feature space with non-specific others may be non-convex
2. may require recombination heuristics
   - $\geq$2 participants may be assigned a unique type

Detecting Participant Types Independently

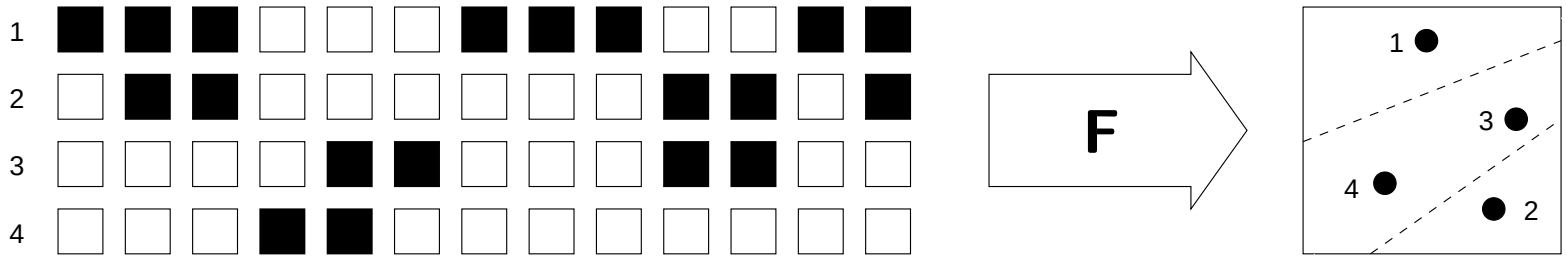


Problems:

1. cannot model interaction with **specific** other types
   - feature space with non-specific others may be non-convex
2. may require recombination heuristics
   - $\geq$2 participants may be assigned a unique type

Solution: model participants **jointly**

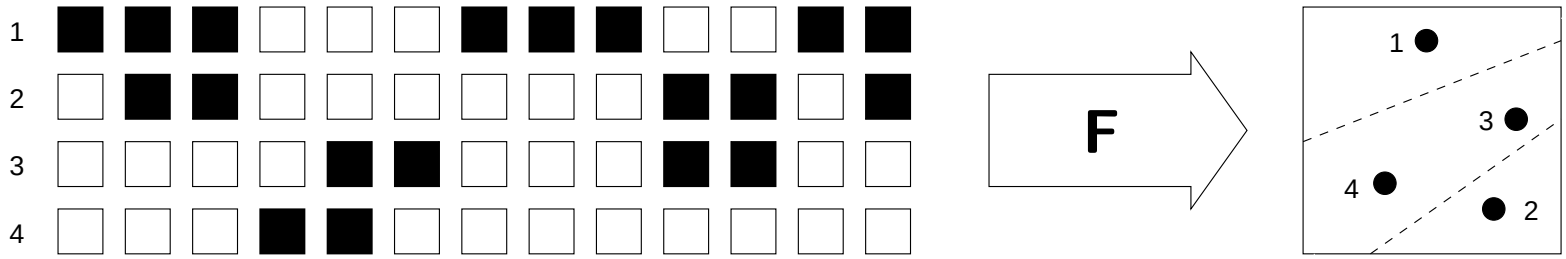
Detecting Participant Types Jointly



1. **F** describes interaction between **all** K participants
2. the most likely **group** assignment $\mathbf{g}^*$ is identified by enumerating over all possible group assignments $\mathbf{g} \in h(\mathcal{C})$

$$\begin{aligned} \mathbf{g}^* &= \arg\max_{\mathbf{g} \in h(\mathcal{C})} P(\mathbf{g} \mid \mathbf{F}) \\ &= \arg\max_{\mathbf{g} \in h(\mathcal{C})} \underbrace{P(\mathbf{g})}_{\text{MM}} \underbrace{P(\mathbf{F} \mid \mathbf{g})}_{\text{BM}} \end{aligned}$$

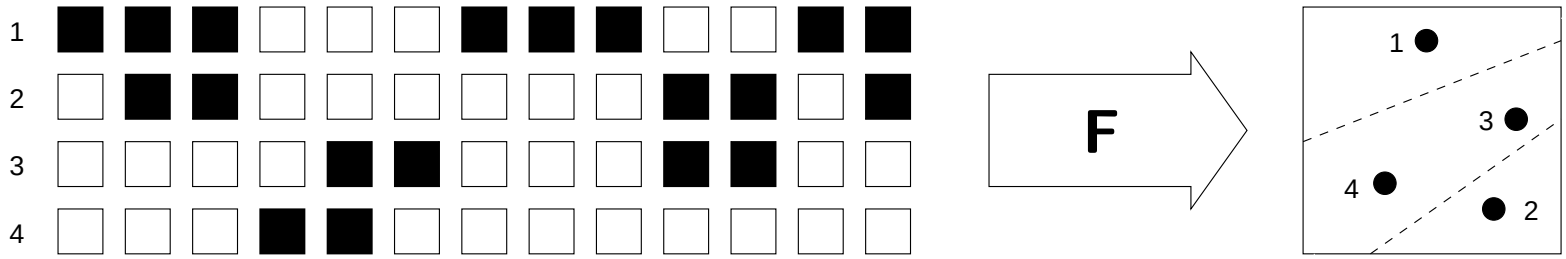
Detecting Participant Types Jointly



1. **F** describes interaction between **all** K participants
2. the most likely **group** assignment $\mathbf{g}^*$ is identified by enumerating over all possible group assignments $\mathbf{g} \in h(\mathcal{C})$

$$\begin{aligned} \mathbf{g}^* &= \arg\max_{\mathbf{g} \in h(\mathcal{C})} P(\mathbf{g} \mid \mathbf{F}) \\ &= \arg\max_{\mathbf{g} \in h(\mathcal{C})} \underbrace{P(\mathbf{g})}_{\text{MM}} \underbrace{P(\mathbf{F} \mid \mathbf{g})}_{\text{BM}} \end{aligned}$$

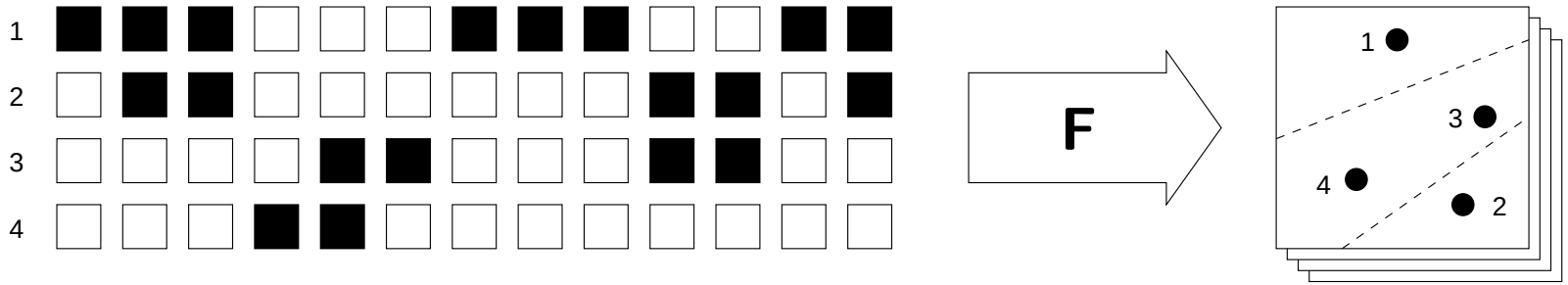
Detecting Participant Types Jointly



1. **F** describes interaction between **all** K participants
2. the most likely **group** assignment $\mathbf{g}^*$ is identified by enumerating over all possible group assignments $\mathbf{g} \in h(\mathcal{C})$

$$\begin{aligned} \mathbf{g}^* &= \arg\max_{\mathbf{g} \in h(\mathcal{C})} P(\mathbf{g} \mid \mathbf{F}) \\ &= \arg\max_{\mathbf{g} \in h(\mathcal{C})} \underbrace{P(\mathbf{g})}_{\text{MM}} \underbrace{P(\mathbf{F} \mid \mathbf{g})}_{\text{BM}} \end{aligned}$$

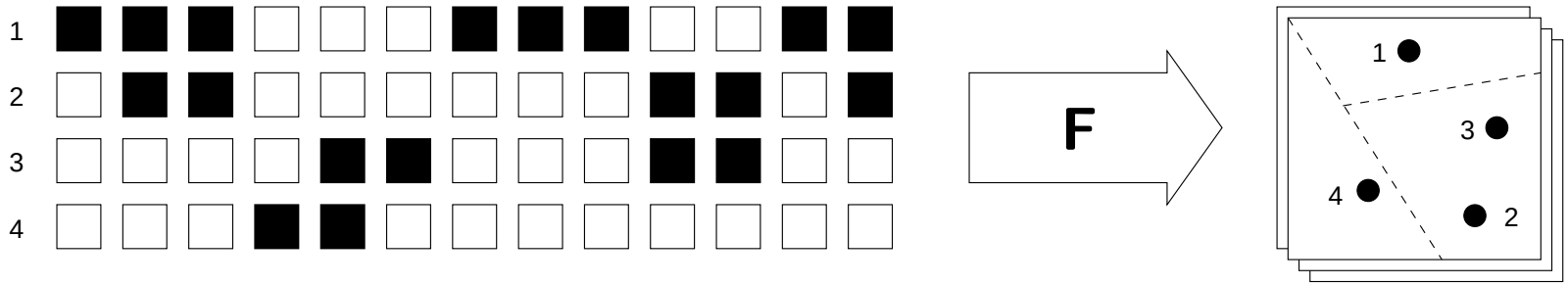
Detecting Participant Types Jointly



1. **F** describes interaction between **all** K participants
2. the most likely **group** assignment $\mathbf{g}^*$ is identified by enumerating over all possible group assignments $\mathbf{g} \in h(\mathcal{C})$

$$\begin{aligned} \mathbf{g}^* &= \arg\max_{\mathbf{g} \in h(\mathcal{C})} P(\mathbf{g} \mid \mathbf{F}) \\ &= \arg\max_{\mathbf{g} \in h(\mathcal{C})} \underbrace{P(\mathbf{g})}_{\text{MM}} \underbrace{P(\mathbf{F} \mid \mathbf{g})}_{\text{BM}} \end{aligned}$$

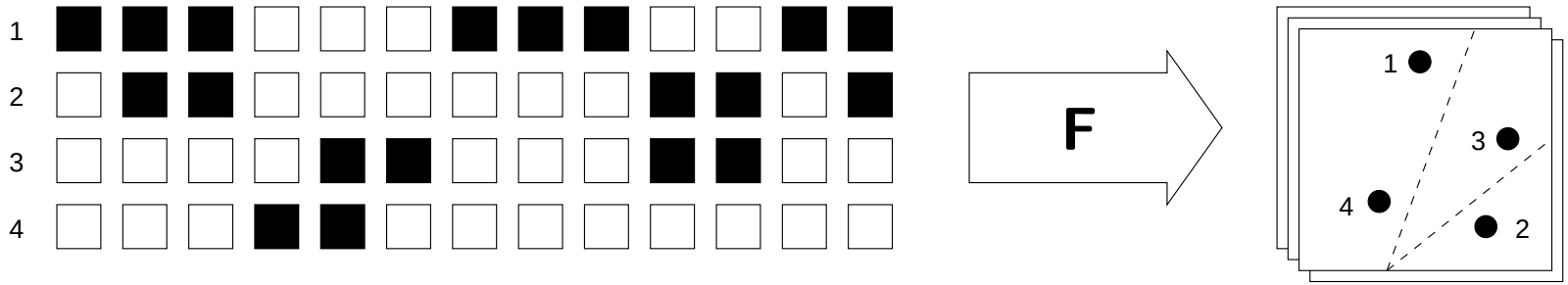
Detecting Participant Types Jointly



1. **F** describes interaction between **all** K participants
2. the most likely **group** assignment $\mathbf{g}^*$ is identified by enumerating over all possible group assignments $\mathbf{g} \in h(\mathcal{C})$

$$\begin{aligned} \mathbf{g}^* &= \arg\max_{\mathbf{g} \in h(\mathcal{C})} P(\mathbf{g} \mid \mathbf{F}) \\ &= \arg\max_{\mathbf{g} \in h(\mathcal{C})} \underbrace{P(\mathbf{g})}_{\text{MM}} \underbrace{P(\mathbf{F} \mid \mathbf{g})}_{\text{BM}} \end{aligned}$$

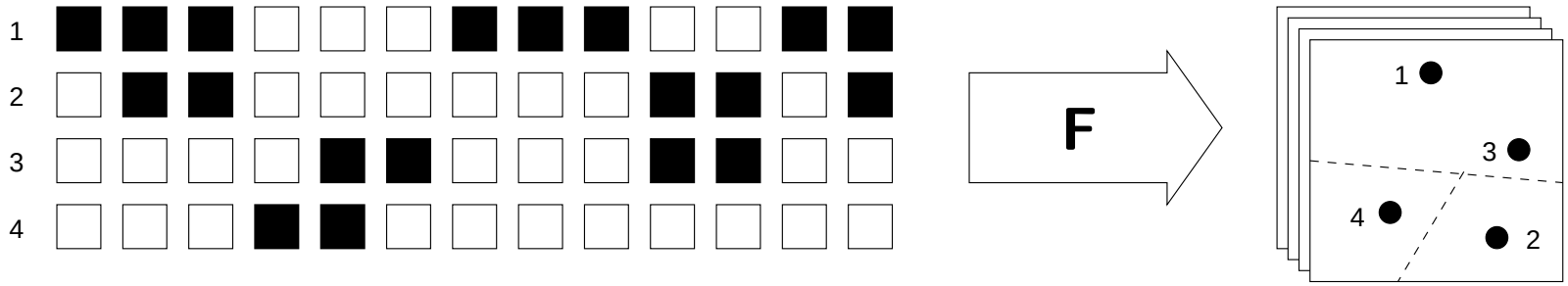
Detecting Participant Types Jointly



1. **F** describes interaction between **all** K participants
2. the most likely **group** assignment $\mathbf{g}^*$ is identified by enumerating over all possible group assignments $\mathbf{g} \in h(\mathcal{C})$

$$\begin{aligned} \mathbf{g}^* &= \arg\max_{\mathbf{g} \in h(\mathcal{C})} P(\mathbf{g} \mid \mathbf{F}) \\ &= \arg\max_{\mathbf{g} \in h(\mathcal{C})} \underbrace{P(\mathbf{g})}_{\text{MM}} \underbrace{P(\mathbf{F} \mid \mathbf{g})}_{\text{BM}} \end{aligned}$$

Detecting Participant Types Jointly



1. **F** describes interaction between **all** K participants
2. the most likely **group** assignment $\mathbf{g}^*$ is identified by enumerating over all possible group assignments $\mathbf{g} \in h(\mathcal{C})$

$$\begin{aligned} \mathbf{g}^* &= \arg\max_{\mathbf{g} \in h(\mathcal{C})} P(\mathbf{g} \mid \mathbf{F}) \\ &= \arg\max_{\mathbf{g} \in h(\mathcal{C})} \underbrace{P(\mathbf{g})}_{\text{MM}} \underbrace{P(\mathbf{F} \mid \mathbf{g})}_{\text{BM}} \end{aligned}$$

What is $h(\mathcal{C})$?

That depends on what $\mathcal{C}$ is...

- each of K types assigned to exactly one participant
- $h(\mathcal{C})$ is a permutation space
- $|h(\mathcal{C})| = K!$

- each participant can be one of any N types
- $h(\mathcal{C})$ is a Cartesian product
- $|h(\mathcal{C})| = N^K$

What is $h(\mathcal{C})$?

Unique Types

- each of K types assigned to exactly one participant
- $h(\mathcal{C})$ is a permutation space
- $|h(\mathcal{C})| = K!$

Non-Unique Types

- each participant can be one of any N types
- $h(\mathcal{C})$ is a Cartesian product
- $|h(\mathcal{C})| = N^K$

What is $h(\mathcal{C})$?

Unique Types

$$\mathcal{C} = \{C_1, C_2, \cdots, C_K\}$$

- each of K types assigned to exactly one participant
- $h(\mathcal{C})$ is a permutation space
- $|h(\mathcal{C})| = K!$

Non-Unique Types

- each participant can be one of any N types
- $h(\mathcal{C})$ is a Cartesian product
- $|h(\mathcal{C})| = N^K$

What is $h(\mathcal{C})$?

Unique Types

$$\mathcal{C} = \{C_1, C_2, \cdots, C_K\}$$

- each of K types assigned to exactly one participant
- $h(\mathcal{C})$ is a permutation space
- $|h(\mathcal{C})| = K!$

Non-Unique Types

- each participant can be one of any N types
- $h(\mathcal{C})$ is a Cartesian product
- $|h(\mathcal{C})| = N^K$

What is $h(\mathcal{C})$?

Unique Types

$$\mathcal{C} = \{C_1, C_2, \cdots, C_K\}$$

- each of K types assigned to exactly one participant
- $h(\mathcal{C})$ is a permutation space
- $|h(\mathcal{C})| = K!$

Non-Unique Types

- each participant can be one of any N types
- $h(\mathcal{C})$ is a Cartesian product
- $|h(\mathcal{C})| = N^K$

What is $h(\mathcal{C})$?

Unique Types

$$\mathcal{C} = \{C_1, C_2, \cdots, C_K\}$$

- each of K types assigned to exactly one participant
- $h(\mathcal{C})$ is a permutation space
- $|h(\mathcal{C})| = K!$

Non-Unique Types

- each participant can be one of any N types
- $h(\mathcal{C})$ is a Cartesian product
- $|h(\mathcal{C})| = N^K$

What is $h(\mathcal{C})$?

Unique Types

$$\mathcal{C} = \{C_1, C_2, \cdots, C_K\}$$

- each of K types assigned to exactly one participant
- $h(\mathcal{C})$ is a permutation space
- $|h(\mathcal{C})| = K!$

Non-Unique Types

$$\mathcal{C} = \{C_1, C_2, \cdots, C_N\}$$

- each participant can be one of any N types
- $h(\mathcal{C})$ is a Cartesian product
- $|h(\mathcal{C})| = N^K$

What is $h(\mathcal{C})$?

Unique Types

$$\mathcal{C} = \{C_1, C_2, \cdots, C_K\}$$

- each of K types assigned to exactly one participant
- $h(\mathcal{C})$ is a permutation space
- $|h(\mathcal{C})| = K!$

Non-Unique Types

$$\mathcal{C} = \{C_1, C_2, \cdots, C_N\}$$

- each participant can be one of any N types
- $h(\mathcal{C})$ is a Cartesian product
- $|h(\mathcal{C})| = N^K$

What is $h(\mathcal{C})$?

Unique Types

$$\mathcal{C} = \{C_1, C_2, \cdots, C_K\}$$

- each of K types assigned to exactly one participant
- $h(\mathcal{C})$ is a permutation space
- $|h(\mathcal{C})| = K!$

Non-Unique Types

$$\mathcal{C} = \{C_1, C_2, \cdots, C_N\}$$

- each participant can be one of any N types
- $h(\mathcal{C})$ is a Cartesian product
- $|h(\mathcal{C})| = N^K$

What is $h(\mathcal{C})$?

Unique Types

$$\mathcal{C} = \{C_1, C_2, \cdots, C_K\}$$

- each of K types assigned to exactly one participant
- $h(\mathcal{C})$ is a permutation space
- $|h(\mathcal{C})| = K!$

Non-Unique Types

$$\mathcal{C} = \{C_1, C_2, \cdots, C_N\}$$

- each participant can be one of any N types
- $h(\mathcal{C})$ is a Cartesian product
- $|h(\mathcal{C})| = N^K$

In our experiments...

- Unique Roles, $\mathcal{C} = \mathcal{R}$
- AMI Meeting Corpus
  - design scenario
  - train: 98 meetings
  - dev: 20 meetings
  - eval: 20 meetings
  - $K = 4$, always
- $\mathcal{R} = \{\mathrm{PM}, \mathrm{ME}, \mathrm{UI}, \mathrm{ID}\}$

- Seniority Levels, $\mathcal{C} = \mathcal{S}$
- ICSI Meeting Corpus
  - naturally occurring
  - 3 meeting types (Bed,Bmr,Bro)
  - train: 33 meetings
  - dev: 18 meetings
  - eval: 16 meetings
  - $3 \geq K \geq 9$
- $\mathcal{S} = \{\mathrm{GRAD}, \mathrm{PHD}, \mathrm{PROF}\}$

In our experiments...

Unique Types

- Unique Roles, $\mathcal{C} = \mathcal{R}$
- AMI Meeting Corpus
  - design scenario
  - train: 98 meetings
  - dev: 20 meetings
  - eval: 20 meetings
  - $K = 4$, always
- $\mathcal{R} = \{\text{PM}, \text{ME}, \text{UI}, \text{ID}\}$

Non-Unique Types

- Seniority Levels, $\mathcal{C} = \mathcal{S}$
- ICSI Meeting Corpus
  - naturally occurring
  - 3 meeting types (Bed,Bmr,Bro)
  - train: 33 meetings
  - dev: 18 meetings
  - eval: 16 meetings
  - $3 \geq K \geq 9$
- $\mathcal{S} = \{\text{GRAD}, \text{PHD}, \text{PROF}\}$

In our experiments...

Unique Types

- Unique Roles, $\mathcal{C} = \mathcal{R}$
- AMI Meeting Corpus
  - design scenario
  - train: 98 meetings
  - dev: 20 meetings
  - eval: 20 meetings
  - $K = 4$, always
- $\mathcal{R} = \{\text{PM}, \text{ME}, \text{UI}, \text{ID}\}$

Non-Unique Types

- Seniority Levels, $\mathcal{C} = \mathcal{S}$
- ICSI Meeting Corpus
  - naturally occurring
  - 3 meeting types (Bed,Bmr,Bro)
  - train: 33 meetings
  - dev: 18 meetings
  - eval: 16 meetings
  - $3 \geq K \geq 9$
- $\mathcal{S} = \{\text{GRAD}, \text{PHD}, \text{PROF}\}$

In our experiments...

Unique Types

- Unique Roles, $\mathcal{C} = \mathcal{R}$
- AMI Meeting Corpus
  - design scenario
  - train: 98 meetings
  - dev: 20 meetings
  - eval: 20 meetings
  - $K = 4$, always
- $\mathcal{R} = \{\text{PM}, \text{ME}, \text{UI}, \text{ID}\}$

Non-Unique Types

- Seniority Levels, $\mathcal{C} = \mathcal{S}$
- ICSI Meeting Corpus
  - naturally occurring
  - 3 meeting types (Bed,Bmr,Bro)
  - train: 33 meetings
  - dev: 18 meetings
  - eval: 16 meetings
  - $3 \geq K \geq 9$
- $\mathcal{S} = \{\text{GRAD}, \text{PHD}, \text{PROF}\}$

In our experiments...

Unique Types

- Unique Roles, $\mathcal{C} = \mathcal{R}$
- AMI Meeting Corpus
  - design scenario
  - train: 98 meetings
  - dev: 20 meetings
  - eval: 20 meetings
  - $K = 4$, always
- $\mathcal{R} = \{\text{PM}, \text{ME}, \text{UI}, \text{ID}\}$

Non-Unique Types

- Seniority Levels, $\mathcal{C} = \mathcal{S}$
- ICSI Meeting Corpus
  - naturally occurring
  - 3 meeting types (Bed,Bmr,Bro)
  - train: 33 meetings
  - dev: 18 meetings
  - eval: 16 meetings
  - $3 \geq K \geq 9$
- $\mathcal{S} = \{\text{GRAD}, \text{PHD}, \text{PROF}\}$

In our experiments...

Unique Types

- Unique Roles, $\mathcal{C} = \mathcal{R}$
- AMI Meeting Corpus
  - design scenario
  - train: 98 meetings
  - dev: 20 meetings
  - eval: 20 meetings
  - $K = 4$, always
- $\mathcal{R} = \{\text{PM}, \text{ME}, \text{UI}, \text{ID}\}$

Non-Unique Types

- Seniority Levels, $\mathcal{C} = \mathcal{S}$
- ICSI Meeting Corpus
  - naturally occurring
  - 3 meeting types (Bed,Bmr,Bro)
  - train: 33 meetings
  - dev: 18 meetings
  - eval: 16 meetings
  - $3 \geq K \geq 9$
- $\mathcal{S} = \{\text{GRAD}, \text{PHD}, \text{PROF}\}$

In our experiments...

Unique Types

- Unique Roles, $\mathcal{C} = \mathcal{R}$
- AMI Meeting Corpus
  - design scenario
  - train: 98 meetings
  - dev: 20 meetings
  - eval: 20 meetings
  - $K = 4$, always
- $\mathcal{R} = \{\text{PM}, \text{ME}, \text{UI}, \text{ID}\}$

Non-Unique Types

- Seniority Levels, $\mathcal{C} = \mathcal{S}$
- ICSI Meeting Corpus
  - naturally occurring
  - 3 meeting types (`Bed`,`Bmr`,`Bro`)
  - train: 33 meetings
  - dev: 18 meetings
  - eval: 16 meetings
  - $3 \geq K \geq 9$
- $\mathcal{S} = \{\text{GRAD}, \text{PHD}, \text{PROF}\}$

In our experiments...

Unique Types

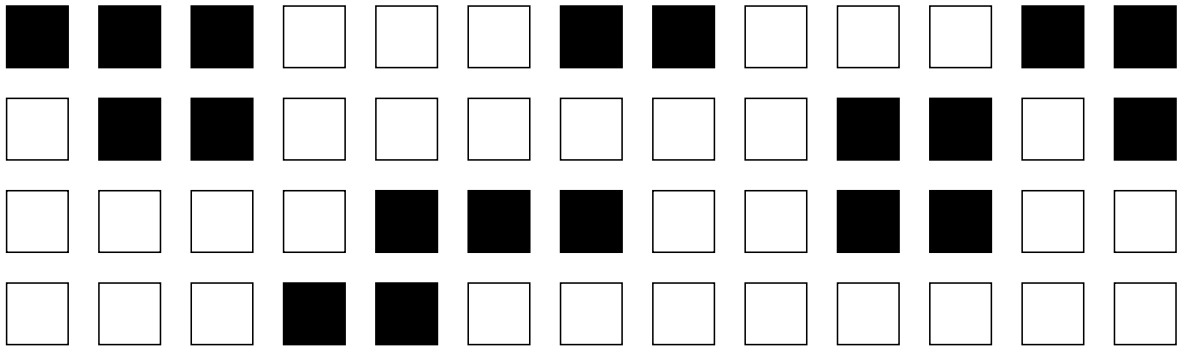
- Unique Roles, $\mathcal{C} = \mathcal{R}$
- AMI Meeting Corpus
  - design scenario
  - train: 98 meetings
  - dev: 20 meetings
  - eval: 20 meetings
  - $K = 4$, always
- $\mathcal{R} = \{\text{PM}, \text{ME}, \text{UI}, \text{ID}\}$

Non-Unique Types

- Seniority Levels, $\mathcal{C} = \mathcal{S}$
- ICSI Meeting Corpus
  - naturally occurring
  - 3 meeting types (`Bed`,`Bmr`,`Bro`)
  - train: 33 meetings
  - dev: 18 meetings
  - eval: 16 meetings
  - $3 \geq K \geq 9$
- $\mathcal{S} = \{\text{GRAD}, \text{PHD}, \text{PROF}\}$

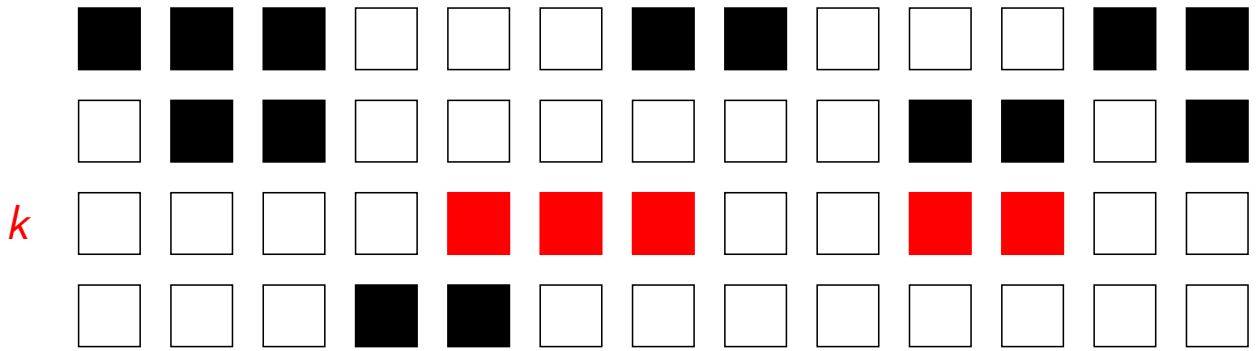
Feature Types in **F**

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**

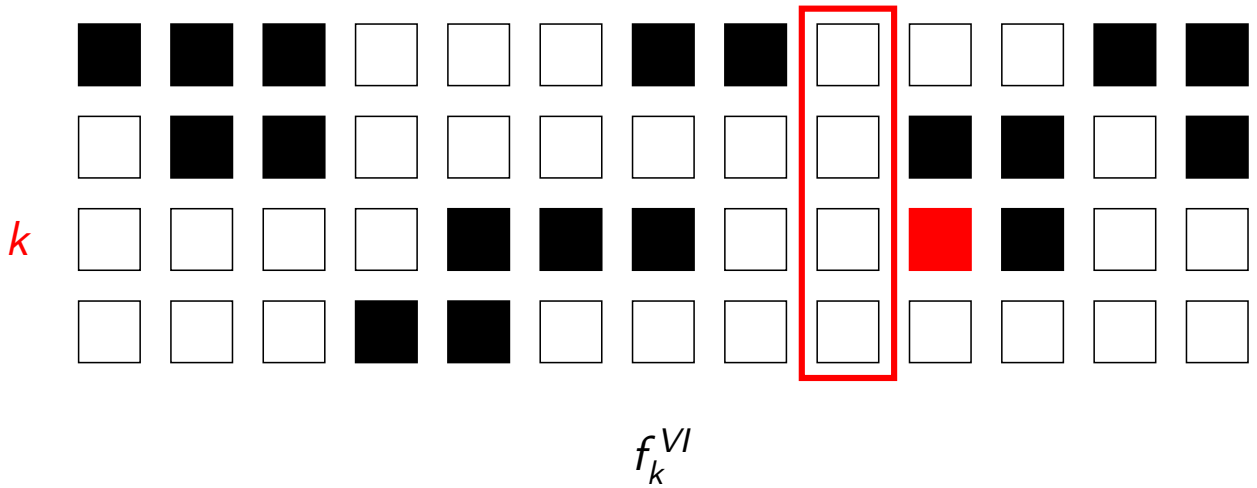
1. probability of vocalizing (V)
2. probability of initiating vocalization (VI) in prior silence
3. probability of continuing vocalization (VC) in prior non-overlap
4. probability of initiating overlap (OI) in prior non-overlap
5. probability of continuing overlap (OC) in prior overlap



f_k^V

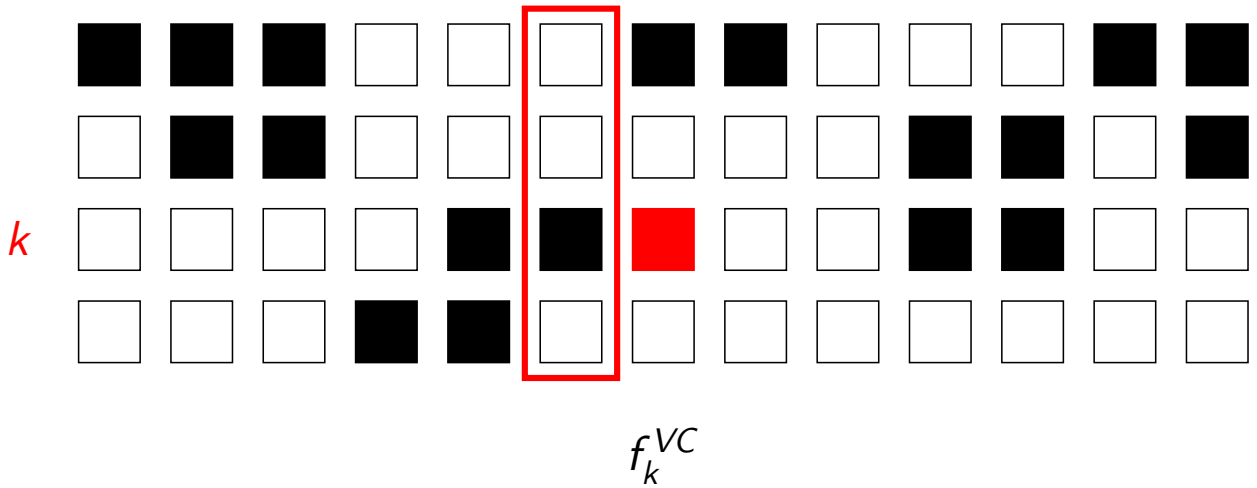
Feature Types in **F**

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



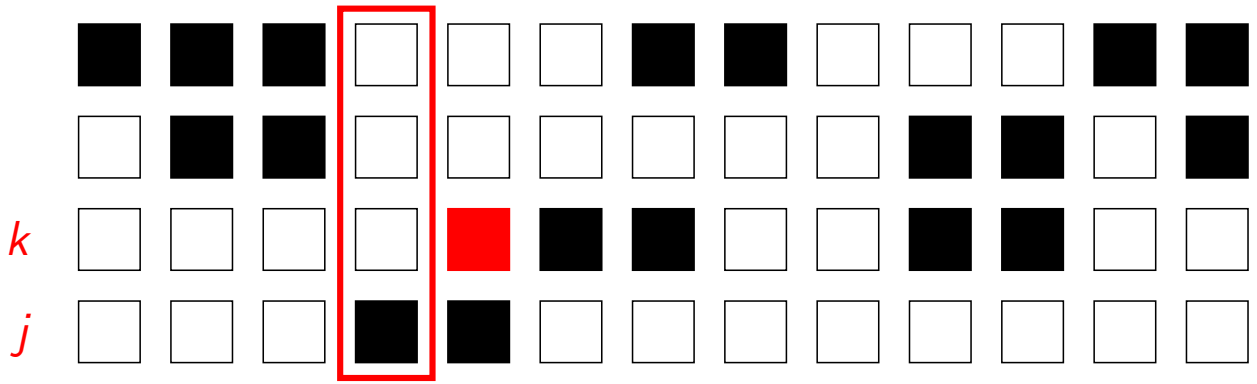
Feature Types in **F**

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**

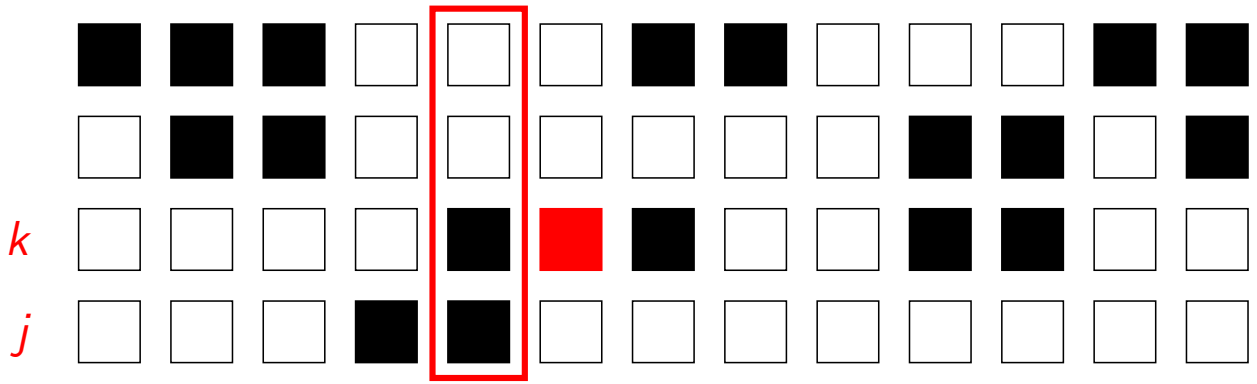
1. probability of vocalizing (V)
2. probability of initiating vocalization (VI) in prior silence
3. probability of continuing vocalization (VC) in prior non-overlap
4. probability of initiating overlap (OI) in prior non-overlap
5. probability of continuing overlap (OC) in prior overlap



$f^{OI}_{k,j}$

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



$f^{OC}_{k,j}$

Models

- behavior model (BM), where θ is a 1-dimensional Gaussian

$$P(\mathbf{F} \mid \mathbf{g}) = \prod_{k=1}^{K} P\left(f_k^V \mid \theta^V_{\mathbf{g}[k]}\right) P\left(f_k^{VI} \mid \theta^{VI}_{\mathbf{g}[k]}\right) P\left(f_k^{VC} \mid \theta^{VC}_{\mathbf{g}[k]}\right) \times \prod_{j \neq k}^{K} P\left(f_{k,j}^{OI} \mid \theta^{OI}_{\mathbf{g}[k],\mathbf{g}[j]}\right) P\left(f_{k,j}^{OC} \mid \theta^{OC}_{\mathbf{g}[k],\mathbf{g}[j]}\right)$$

- membership model (MM)

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

Models

- behavior model (BM), where θ is a 1-dimensional Gaussian

$$P(\mathbf{F}\,|\,\mathbf{g}) = \prod_{k=1}^{K} P\left(f_k^V \,|\, \theta^V_{\mathbf{g}[k]}\right) P\left(f_k^{VI} \,|\, \theta^{VI}_{\mathbf{g}[k]}\right) P\left(f_k^{VC} \,|\, \theta^{VC}_{\mathbf{g}[k]}\right) \times \prod_{j\neq k}^{K} P\left(f_{k,j}^{OI} \,|\, \theta^{OI}_{\mathbf{g}[k],\mathbf{g}[j]}\right) P\left(f_{k,j}^{OC} \,|\, \theta^{OC}_{\mathbf{g}[k],\mathbf{g}[j]}\right)$$

- membership model (MM)

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

Models

- behavior model (BM), where θ is a 1-dimensional Gaussian

$$P(\mathbf{F}\,|\,\mathbf{g}) = \prod_{k=1}^{K} P\left(f_k^V \,|\, \theta^V_{\mathbf{g}[k]}\right) P\left(f_k^{VI} \,|\, \theta^{VI}_{\mathbf{g}[k]}\right) P\left(f_k^{VC} \,|\, \theta^{VC}_{\mathbf{g}[k]}\right) \times \prod_{j\neq k}^{K} P\left(f_{k,j}^{OI} \,|\, \theta^{OI}_{\mathbf{g}[k],\mathbf{g}[j]}\right) P\left(f_{k,j}^{OC} \,|\, \theta^{OC}_{\mathbf{g}[k],\mathbf{g}[j]}\right)$$

- membership model (MM)

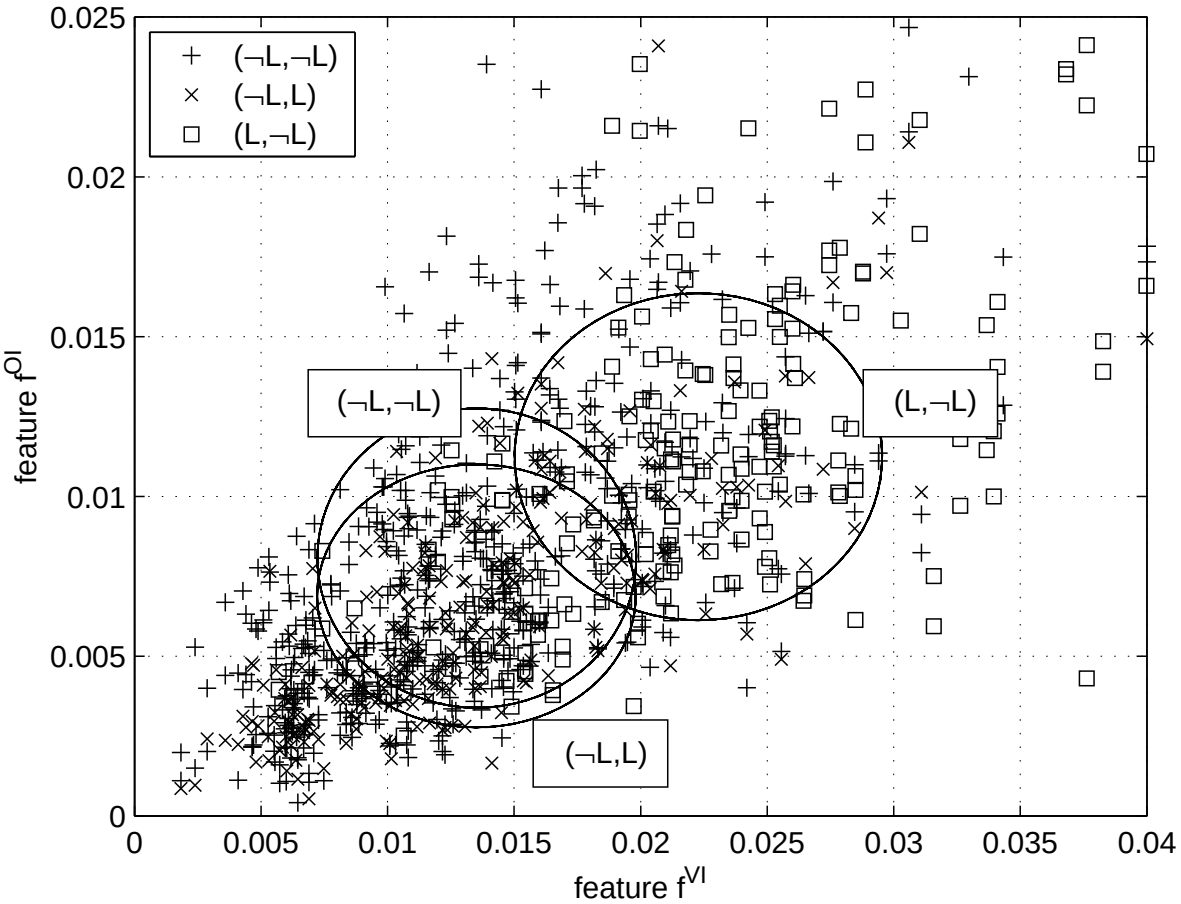
$$P(\mathbf{g}) = \frac{1}{Z_{\mathbf{g}}} \prod_{k=1}^{K} P(\mathbf{g}[k])$$

Unique Role $\mathcal{R}$ Classification

Feature Type	AMI $\mathcal{R}$
f_k^V	44
f_k^{VI}	*41
f_k^{VC}	34
$f_{k,j}^{OI}$	*53
$f_{k,j}^{OC}$	—
best*	53
all	46
priors	25

Aside: Looking for the Leader

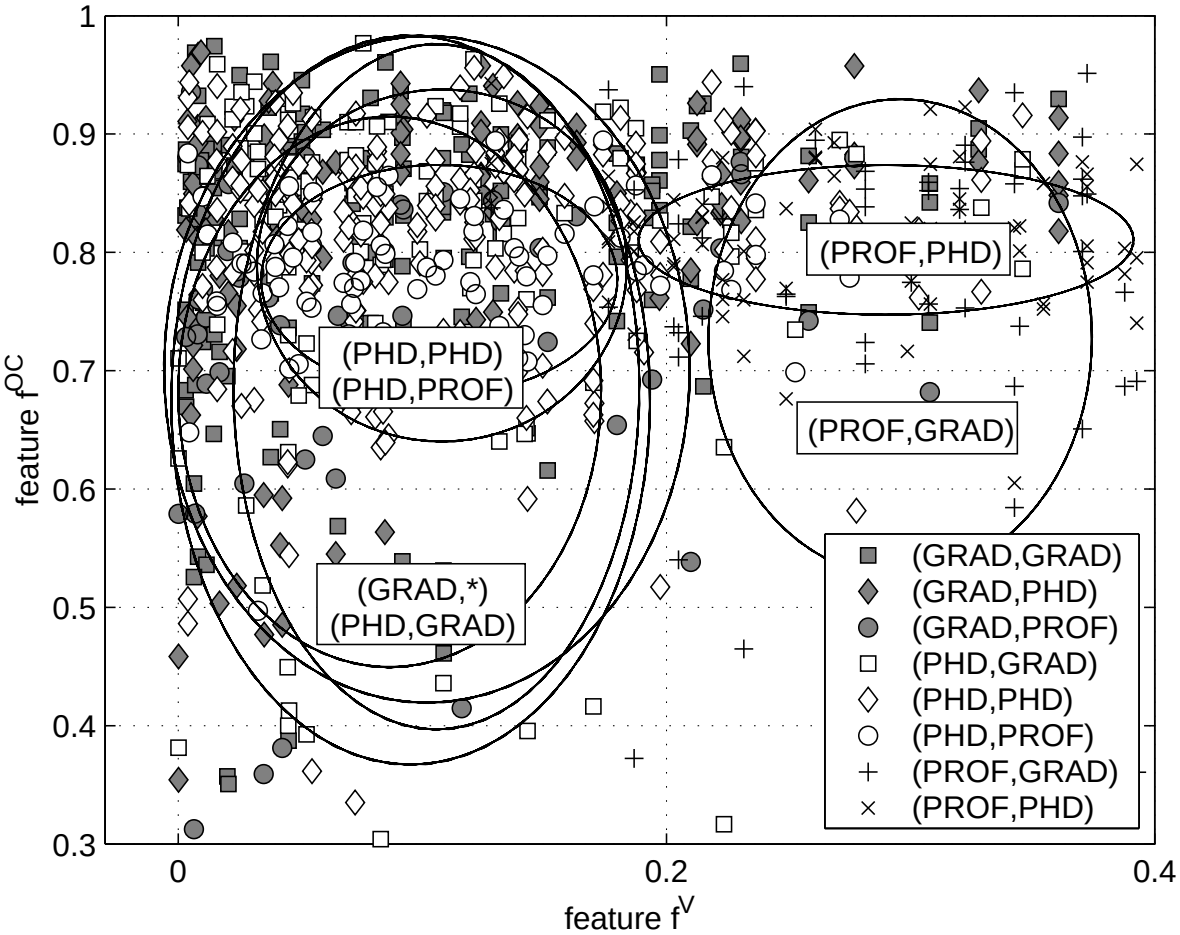
- find one unique role only, $\mathbf{g}[k] \in \mathcal{L} = \{L \equiv \text{PM}, \neg L\}$



Leader $\mathcal{L}$ Detection

Feature Type	AMI	
	$\mathcal{R}$	$\mathcal{L}$
f_k^V	44	—
f_k^{VI}	*41	*60
f_k^{VC}	34	—
$f_{k,j}^{OI}$	*53	*60
$f_{k,j}^{OC}$	—	—
best*	53	60
all	46	75
priors	25	25

Seniority Level Feature Distributions



Seniority Level $\mathcal{S}$ Classification

Feature Type	AMI		ICSI
	$\mathcal{R}$	$\mathcal{L}$	$\mathcal{S}$
f_k^V	44	—	*52
f_k^{VI}	*41	*60	52
f_k^{VC}	34	—	—
$f_{k,j}^{OI}$	*53	*60	*59
$f_{k,j}^{OC}$	—	—	*59
best*	53	60	61
all	46	75	58
priors	25	25	45

Conversation-Type-Dependent $\mathcal{S}$ Classification

- condition models on automatically inferred meeting type

Feature Type	AMI		ICSI	
	$\mathcal{R}$	$\mathcal{L}$	$\mathcal{S}$	$\mathcal{S}\|t^*$
f_k^V	44	—	*52	*57
f_k^{VI}	*41	*60	52	56
f_k^{VC}	34	—	—	62
$f_{k,j}^{OI}$	*53	*60	*59	*59
$f_{k,j}^{OC}$	—	—	*59	*63
best*	53	60	61	67
all	46	75	58	57
priors	25	25	45	45

Conclusions

- assigned unique roles $\mathcal{R}$ in the AMI corpus
  - 53% accuracy, 37% rel error reduction over baseline
  - improves to 75%, when only manager (PM) is sought
  - best features: initiation of talkspurts in silence and in overlap
- seniority level $\mathcal{S}$ in the ICSI corpus

Conclusions

- assigned unique roles $\mathcal{R}$ in the AMI corpus
  - 53% accuracy, 37% rel error reduction over baseline
  - improves to 75%, when only manager (PM) is sought
  - best features: initiation of talkspurts in silence and in overlap
- seniority level $\mathcal{S}$ in the ICSI corpus
  - 61% accuracy, 29% rel error reduction over baseline
  - improves to 67%, with conditioning on inferred meeting type
  - improves to 73%, with conditioning on true meeting type
  - best features: overall talkspurt production, initiation and continuation of talkspurts in overlap

Conclusions

- assigned unique roles $\mathcal{R}$ in the AMI corpus
  - 53% accuracy, 37% rel error reduction over baseline
  - improves to 75%, when only manager (PM) is sought
  - best features: initiation of talkspurts in silence and in overlap
- seniority level $\mathcal{S}$ in the ICSI corpus
  - 61% accuracy, 29% rel error reduction over baseline
  - improves to 67%, with conditioning on inferred meeting type
  - improves to 73%, with conditioning on true meeting type
  - best features: overall talkspurt production, initiation and continuation of talkspurts in overlap

Conclusions

- assigned unique roles $\mathcal{R}$ in the AMI corpus
  - 53% accuracy, 37% rel error reduction over baseline
  - improves to 75%, when only manager (PM) is sought
  - best features: initiation of talkspurts in silence and in overlap
- seniority level $\mathcal{S}$ in the ICSI corpus
  - 61% accuracy, 29% rel error reduction over baseline
  - improves to 67%, with conditioning on inferred meeting type
  - improves to 73%, with conditioning on true meeting type
  - best features: overal talkspurt production, initiation and continuation of talkspurts in overlap

Conclusions

- assigned unique roles $\mathcal{R}$ in the AMI corpus
  - 53% accuracy, 37% rel error reduction over baseline
  - improves to 75%, when only manager (PM) is sought
  - best features: initiation of talkspurts in silence and in overlap
- seniority level $\mathcal{S}$ in the ICSI corpus
  - 61% accuracy, 29% rel error reduction over baseline
  - improves to 67%, with conditioning on inferred meeting type
  - improves to 73%, with conditioning on true meeting type
  - best features: overal talkspurt production, initiation and continuation of talkspurts in overlap

Conclusions

- assigned unique roles $\mathcal{R}$ in the AMI corpus
  - 53% accuracy, 37% rel error reduction over baseline
  - improves to 75%, when only manager (PM) is sought
  - best features: initiation of talkspurts in silence and in overlap
- seniority level $\mathcal{S}$ in the ICSI corpus
  - 61% accuracy, 29% rel error reduction over baseline
  - improves to 67%, with conditioning on inferred meeting type
  - improves to 73%, with conditioning on true meeting type
  - best features: overal talkspurt production, initiation and continuation of talkspurts in overlap

Conclusions

- assigned unique roles $\mathcal{R}$ in the AMI corpus
  - 53% accuracy, 37% rel error reduction over baseline
  - improves to 75%, when only manager (PM) is sought
  - best features: initiation of talkspurts in silence and in overlap
- seniority level $\mathcal{S}$ in the ICSI corpus
  - 61% accuracy, 29% rel error reduction over baseline
  - improves to 67%, with conditioning on inferred meeting type
  - improves to 73%, with conditioning on true meeting type
  - best features: overal talkspurt production, initiation and continuation of talkspurts in overlap

(Potential) Implications

1. **Participant Characterization**
   - talkspurt deployment timing is predictive
   - first baseline for several of the explored tasks
   - proposed framework allows for inclusion of potentially complementary information, to prosodic/lexical/semantic features
2. **Dialogue Systems**
3. **Speech Activity Detection**

(Potential) Implications

1. **Participant Characterization**
   - talkspurt deployment timing is predictive
   - first baseline for several of the explored tasks
   - proposed framework allows for inclusion of potentially complementary information, to prosodic/lexical/semantic features
2. **Dialogue Systems**
3. **Speech Activity Detection**

(Potential) Implications

1. **Participant Characterization**
   - talkspurt deployment timing is predictive
   - first baseline for several of the explored tasks
   - proposed framework allows for inclusion of potentially complementary information, to prosodic/lexical/semantic features
2. **Dialogue Systems**
   - agent talkspurt deployment may contribute to agent personality
3. **Speech Activity Detection**

(Potential) Implications

1. **Participant Characterization**
   - talkspurt deployment timing is predictive
   - first baseline for several of the explored tasks
   - proposed framework allows for inclusion of potentially complementary information, to prosodic/lexical/semantic features
2. **Dialogue Systems**
   - agent talkspurt deployment may contribute to agent personality
3. **Speech Activity Detection**
   - performance likely to improve with conditioning on participant characteristics
   - or joint inference of SAD and participant characteristics

(Potential) Implications

1. **Participant Characterization**
   - talkspurt deployment timing is predictive
   - first baseline for several of the explored tasks
   - proposed framework allows for inclusion of potentially complementary information, to prosodic/lexical/semantic features
2. **Dialogue Systems**
   - agent talkspurt deployment may contribute to agent personality
3. **Speech Activity Detection**
   - performance likely to improve with conditioning on participant characteristics
   - or joint inference of SAD and participant characteristics

(Potential) Implications

1. **Participant Characterization**
   - talkspurt deployment timing is predictive
   - first baseline for several of the explored tasks
   - proposed framework allows for inclusion of potentially complementary information, to prosodic/lexical/semantic features
2. **Dialogue Systems**
   - agent talkspurt deployment may contribute to agent personality
3. **Speech Activity Detection**
   - performance likely to improve with conditioning on participant characteristics
   - or joint inference of SAD and participant characteristics

(Potential) Implications

1. **Participant Characterization**
   - talkspurt deployment timing is predictive
   - first baseline for several of the explored tasks
   - proposed framework allows for inclusion of potentially complementary information, to prosodic/lexical/semantic features
2. **Dialogue Systems**
   - agent talkspurt deployment may contribute to agent personality
3. **Speech Activity Detection**
   - performance likely to improve with conditioning on participant characteristics
   - or joint inference of SAD and participant characteristics

Thank you for attending.

Many thanks also to:

- Jean Carletta, for many helpful comments
- Liz Shriberg, for access to the ICSI MRDA Corpus