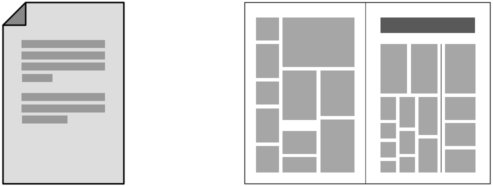
Modeling Norms of Turn-Taking in Multi-Party Conversation

Kornel Laskowski

Carnegie Mellon University
Pittsburgh PA, USA

13 July, 2010

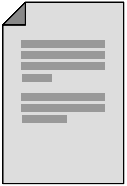
Comparing Written Documents



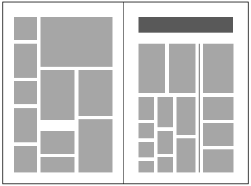
$\mathbf{w}_1$ $\mathbf{w}_2$

- If have a form for a density model Θ of word sequences, and
- techniques for estimating the parameters of Θ from data, and
- techniques for estimating $P(\mathbf{w} \mid \Theta)$,
- Can easily compare $\mathbf{w}_1$ with $\mathbf{w}_2$, with respect to
- how far each deviates from the norms encoded in Θ

Comparing Written Documents



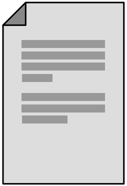
$\mathbf{w}_1$



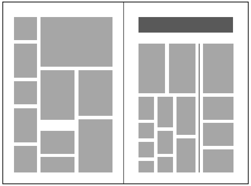
$\mathbf{w}_2$

- If have a form for a density model $\mathbf{\Theta}$ of word sequences, and
- techniques for estimating the parameters of $\mathbf{\Theta}$ from data, and
- techniques for estimating $P(\mathbf{w} \mid \mathbf{\Theta})$,
- Can easily compare $\mathbf{w}_1$ with $\mathbf{w}_2$, with respect to
- how far each deviates from the norms encoded in $\mathbf{\Theta}$

Comparing Written Documents



$\mathbf{w}_1$

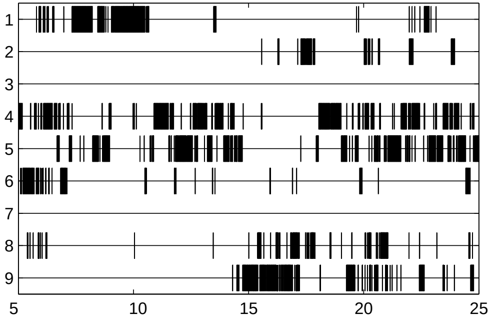


$\mathbf{w}_2$

- If have a form for a density model $\mathbf{\Theta}$ of word sequences, and
- techniques for estimating the parameters of $\mathbf{\Theta}$ from data, and
- techniques for estimating $P(\mathbf{w} \mid \mathbf{\Theta})$,
- Can easily compare $\mathbf{w}_1$ with $\mathbf{w}_2$, with respect to
- how far each deviates from the norms encoded in $\mathbf{\Theta}$

Representing Spoken Documents

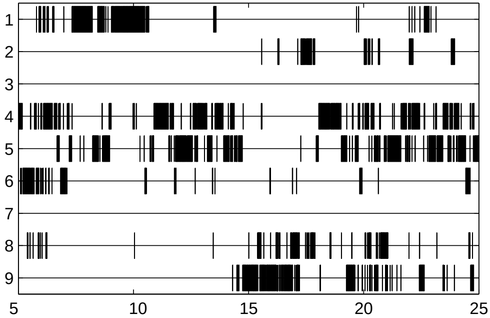
- $K > 1$ sources (participants)



- interaction chronograph (Chapple, 1939)
- aka vocal interaction record (Dabbs & Ruback, 1987)
- a **content-independent** representation

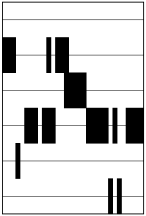
Representing Spoken Documents

- $K > 1$ sources (participants)

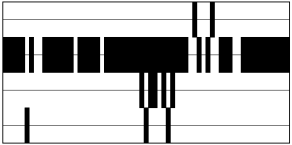


- interaction chronograph (Chapple, 1939)
- aka vocal interaction record (Dabbs & Ruback, 1987)
- a **content-independent** representation

Comparing Spoken Documents



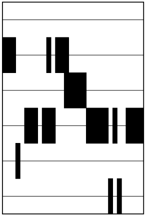
$\mathbf{Q}_1$



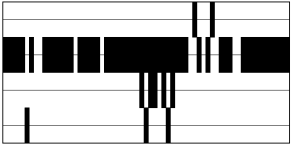
$\mathbf{Q}_2$

- Cannot compare spoken documents $\mathbf{Q}_1$ and $\mathbf{Q}_2$
  - no candidate form for model $\boldsymbol{\Theta}$,
  - no techniques of estimating the parameters of $\boldsymbol{\Theta}$ from data,
  - no techniques of estimating $P(\mathbf{Q} \mid \boldsymbol{\Theta})$.
- Comparison must remain **manual** and **qualitative**.

Comparing Spoken Documents



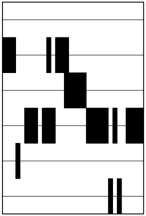
$\mathbf{Q}_1$



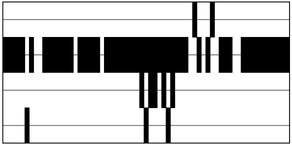
$\mathbf{Q}_2$

- Cannot compare spoken documents $\mathbf{Q}_1$ and $\mathbf{Q}_2$
  - no candidate form for model $\boldsymbol{\Theta}$,
  - no techniques of estimating the parameters of $\boldsymbol{\Theta}$ from data,
  - no techniques of estimating $P(\mathbf{Q} \mid \boldsymbol{\Theta})$.
- Comparison must remain **manual** and **qualitative**.

Comparing Spoken Documents



$\mathbf{Q}_1$



$\mathbf{Q}_2$

- Cannot compare spoken documents $\mathbf{Q}_1$ and $\mathbf{Q}_2$
  - no candidate form for model $\mathbf{\Theta}$,
  - no techniques of estimating the parameters of $\mathbf{\Theta}$ from data,
  - no techniques of estimating $P(\mathbf{Q} \mid \mathbf{\Theta})$.
- Comparison must remain **manual** and **qualitative**.

Wouldn't It Be Nice if We Could ...

- Compare meetings in organizations to determine which interaction patterns correlate with successful business practice?
- Find instants within conversations where interaction management breaks down (hotspots)?
- Classify conversations according to a spectrum of interactivity?
- Contrast conversational behavior across languages and cultures?
- Assess emergent turn-taking performance in dialogue systems?

Outline of this Talk

1. Compositional Modeling Framework
2. Direct estimation in compositional models
3. "Extended Degree of Overlap" (EDO) Model
4. Experiments with Naturally Occurring Conversation
   - within-conversation prediction
   - across-conversation prediction
5. Summary

Turns and Talk Spurts ...

1. **turn-taking**: generally observed phenomenon in conversation
2. but **turn**: ?? (no generally agreed upon definition)

3. here: **turn** $\equiv$ **(talk) spurt** (Norwine & Murphy, 1938)
   - prefer "speech regions uninterrupted by pauses longer than 500 ms" (Shriberg et al, 2001)
   - with a threshold $T_{\square} = 300$ ms (NIST RT Evaluations, 2002–)
   - similar to **inter-pausal unit**, $T_{\square} = 100$ms (Koiso et al, 1998)

4. specific choice may have minor numerical consequences
5. but no impact on the mathematical viability of the modeling techniques of this work

Turns and Talk Spurts ...

1. **turn-taking**: generally observed phenomenon in conversation
2. but **turn**: ?? (no generally agreed upon definition)
3. here: **turn** $\equiv$ **(talk) spurt** (Norwine & Murphy, 1938)
   - prefer "speech regions uninterrupted by pauses longer than 500 ms" (Shriberg et al, 2001)
   - with a threshold $T_{\square} = 300$ ms (NIST RT Evaluations, 2002–)
   - similar to **inter-pausal unit**, $T_{\square} = 100$ms (Koiso et al, 1998)
4. specific choice may have minor numerical consequences
5. but no impact on the mathematical viability of the modeling techniques of this work

Turns and Talk Spurts ...

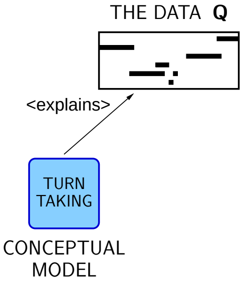
1. **turn-taking**: generally observed phenomenon in conversation
2. but **turn**: ?? (no generally agreed upon definition)
3. here: **turn** $\equiv$ **(talk) spurt** (Norwine & Murphy, 1938)
   - prefer "speech regions uninterrupted by pauses longer than 500 ms" (Shriberg et al, 2001)
   - with a threshold $T_{\square} = 300$ ms (NIST RT Evaluations, 2002–)
   - similar to **inter-pausal unit**, $T_{\square} = 100$ms (Koiso et al, 1998)
4. specific choice may have minor numerical consequences
5. but no impact on the mathematical viability of the modeling techniques of this work

... and Models of Either or Both

THE DATA **Q**

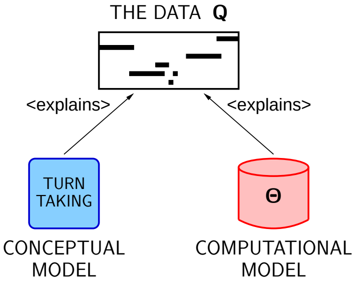
- data **Q**: speech activity in time and across participants
- **turn-taking**: name of a conceptual model
- **goal in this work**: propose a **computational** model

… and Models of Either or Both



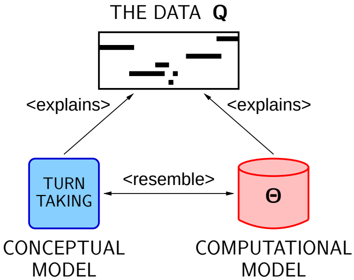
- data **Q**: speech activity in time and across participants
- **turn-taking**: name of a conceptual model
- **goal in this work**: propose a **computational** model

... and Models of Either or Both



- data **Q**: speech activity in time and across participants
- **turn-taking**: name of a conceptual model
- **goal in this work**: propose a **computational** model

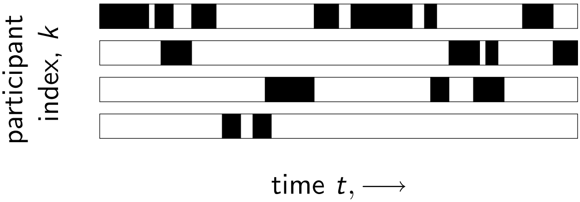
... and Models of Either or Both



- data **Q**: speech activity in time and across participants
- **turn-taking**: name of a conceptual model
- **goal in this work**: propose a **computational** model

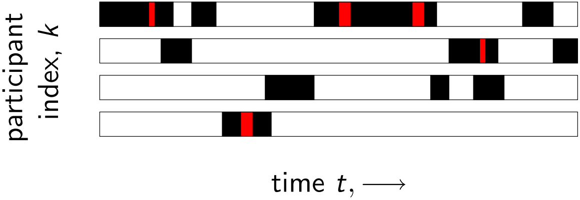
What Is $\mathbf{Q}$?

1. time-align speech $\square$/$\blacksquare$ activity of all K participants
2. close all $\square$ gaps shorter than $T_{\square} = 300$ ms
3. discretize with a frame step of 100 ms
4. yields a $\mathbf{Q} \in \{\square, \blacksquare\}^{K \times T}$



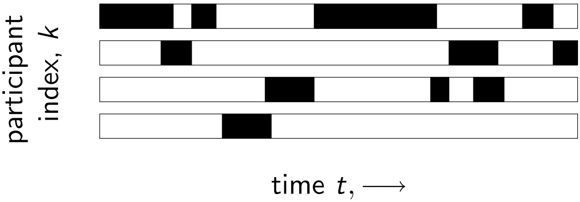
What Is $\mathbf{Q}$?

1. time-align speech $\square/\blacksquare$ activity of all K participants
2. close all $\square$ gaps shorter than $T_\square = 300$ ms
3. discretize with a frame step of 100 ms
4. yields a $\mathbf{Q} \in \{\square, \blacksquare\}^{K \times T}$



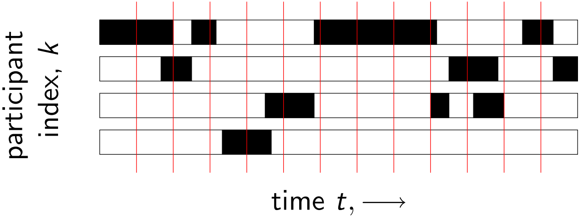
What Is $\mathbf{Q}$?

1. time-align speech $\square/\blacksquare$ activity of all K participants
2. close all $\square$ gaps shorter than $T_{\square} = 300$ ms
3. discretize with a frame step of 100 ms
4. yields a $\mathbf{Q} \in \{\square, \blacksquare\}^{K \times T}$



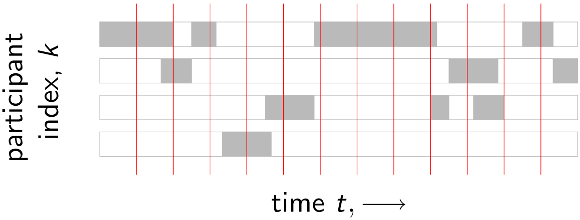
What Is $\mathbf{Q}$?

1. time-align speech $\square/\blacksquare$ activity of all K participants
2. close all $\square$ gaps shorter than $T_{\square} = 300$ ms
3. discretize with a frame step of 100 ms
4. yields a $\mathbf{Q} \in \{\square, \blacksquare\}^{K \times T}$



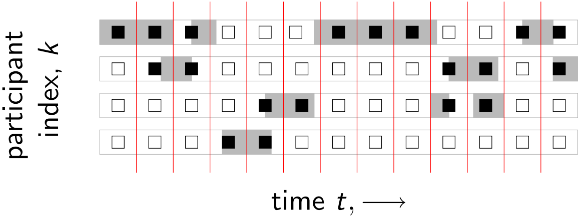
What Is **Q**?

1. time-align speech □/■ activity of all K participants
2. close all □ gaps shorter than $T_{\square} = 300$ ms
3. discretize with a frame step of 100 ms
4. yields a $\mathbf{Q} \in \{\square, \blacksquare\}^{K \times T}$



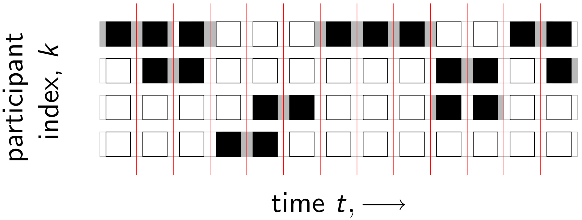
What Is Q?

1. time-align speech □/■ activity of all K participants
2. close all □ gaps shorter than $T_{\square} = 300$ ms
3. discretize with a frame step of 100 ms
4. yields a $\mathbf{Q} \in \{\square, \blacksquare\}^{K \times T}$



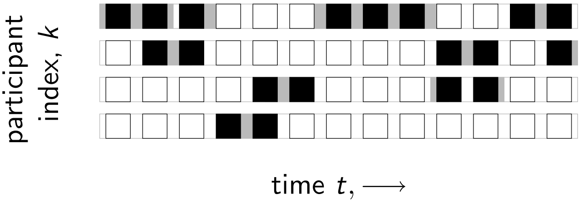
What Is $\mathbf{Q}$?

1. time-align speech $\square/\blacksquare$ activity of all K participants
2. close all $\square$ gaps shorter than $T_{\square} = 300$ ms
3. discretize with a frame step of 100 ms
4. yields a $\mathbf{Q} \in \{\square, \blacksquare\}^{K \times T}$



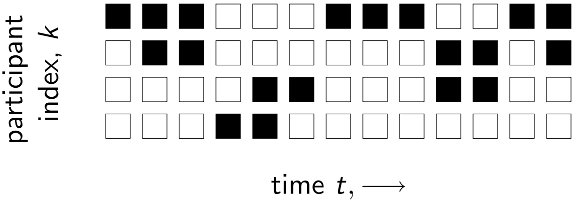
What Is $\mathbf{Q}$?

1. time-align speech $\square/\blacksquare$ activity of all K participants
2. close all $\square$ gaps shorter than $T_{\square} = 300$ ms
3. discretize with a frame step of 100 ms
4. yields a $\mathbf{Q} \in \{\square, \blacksquare\}^{K \times T}$



What Is **Q**?

1. time-align speech $\square/\blacksquare$ activity of all K participants
2. close all $\square$ gaps shorter than $T_{\square} = 300$ ms
3. discretize with a frame step of 100 ms
4. yields a $\mathbf{Q} \in \{\square, \blacksquare\}^{K \times T}$



Modeling A Vector-Valued Markov Process

- model conversation as a Markov process

$$\cdots, \quad \mathbf{q}_{t-1} = \begin{bmatrix} \square \\ \square \\ \blacksquare \\ \square \end{bmatrix}, \quad \mathbf{q}_{t} = \begin{bmatrix} \blacksquare \\ \square \\ \square \\ \square \end{bmatrix}, \quad \mathbf{q}_{t+1} = \begin{bmatrix} \blacksquare \\ \square \\ \square \\ \square \end{bmatrix}, \quad \cdots$$

- then (assuming first-order model $\mathbf{\Theta}$)

- easy to train, just count bigrams !

Modeling A Vector-Valued Markov Process

- model conversation as a Markov process

$$\cdots, \quad \mathbf{q}_{t-1} = \begin{bmatrix} \square \\ \square \\ \blacksquare \\ \square \end{bmatrix}, \quad \mathbf{q}_{t} = \begin{bmatrix} \blacksquare \\ \square \\ \square \\ \square \end{bmatrix}, \quad \mathbf{q}_{t+1} = \begin{bmatrix} \blacksquare \\ \square \\ \square \\ \square \end{bmatrix}, \quad \cdots$$

- then (assuming first-order model $\mathbf{\Theta}$)

$$\begin{aligned} P\left(\mathbf{Q} \mid \mathbf{\Theta}\right) &= P_0 \prod_{t=1}^{T} P\left(\mathbf{q}_t \mid \mathbf{q}_0, \mathbf{q}_1, \cdots, \mathbf{q}_{t-1}, \mathbf{\Theta}\right) \\ &\doteq P_0 \prod_{t=1}^{T} P\left(\mathbf{q}_t \mid \mathbf{q}_{t-1}, \mathbf{\Theta}\right) \end{aligned}$$

- easy to train, just count bigrams !

Modeling A Vector-Valued Markov Process

- model conversation as a Markov process

$$\cdots, \quad \mathbf{q}_{t-1} = \begin{bmatrix} \square \\ \square \\ \blacksquare \\ \square \end{bmatrix}, \quad \mathbf{q}_{t} = \begin{bmatrix} \blacksquare \\ \square \\ \square \\ \square \end{bmatrix}, \quad \mathbf{q}_{t+1} = \begin{bmatrix} \blacksquare \\ \square \\ \square \\ \square \end{bmatrix}, \quad \cdots$$

- then (assuming first-order model $\mathbf{\Theta}$)

$$\begin{aligned} P(\mathbf{Q} \mid \mathbf{\Theta}) &= P_0 \prod_{t=1}^{T} P(\mathbf{q}_t \mid \mathbf{q}_0, \mathbf{q}_1, \cdots, \mathbf{q}_{t-1}, \mathbf{\Theta}) \\ &\doteq P_0 \prod_{t=1}^{T} P(\mathbf{q}_t \mid \mathbf{q}_{t-1}, \mathbf{\Theta}) \end{aligned}$$

- easy to train, just count bigrams !

Some Past Work

- interaction chronography (Chapple, 1939; Chapple, 1949)
- **modeling in dialogue**: $K = 2$
  - telecomminications (Norwine & Murphy, 1938; Brady, 1969)
  - sociolinguistics (Jaffe & Feldstein, 1970)
  - psycholinguistics (Dabbs & Ruback, 1987)
  - dialogue systems (cf. Raux, 2008)
- **modeling in multi-party settings**: $K > 2$
  - psycholinguistics: GroupTalk model (Dabbs et al, 1987)
    - not quite serviceable for current task
  - pre-ASR segmentation: EDO model (Laskowski & Schultz, 2007)

Defining Turn-Taking Perplexity (PPL)

In language modeling,

$\mathbf{w}$: word $\|\mathbf{w}\|$-sequence

$\mathbf{\Theta}$: "language model"

$$\mathrm{NLL} = -\frac{1}{\|\mathbf{w}\|} \log_e P(\mathbf{w} \mid \mathbf{\Theta})$$

$$\mathrm{PPL} = 10^{\mathrm{NLL}}$$

Here,

$\mathbf{Q}$: $K \times T$ chronograph

$\mathbf{\Theta}$: "turn-taking model"

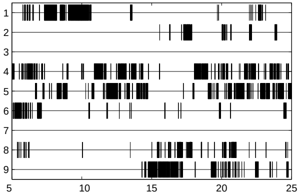
$$\mathrm{NLL} = -\frac{1}{KT} \log_2 P(\mathbf{Q} \mid \mathbf{\Theta})$$

$$\mathrm{PPL} = 2^{\mathrm{NLL}} = \left(P(\mathbf{Q} \mid \mathbf{\Theta})\right)^{-1/KT}$$

- Can also window negative log-likelihood (NLL) to yield a measure of local perplexity (in time).

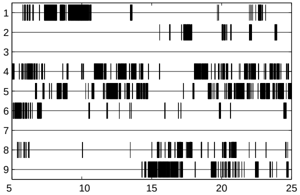
An Example of the Perplexity Trajectory in Time

1. obtain $\mathbf{Q}$ for ICSI `Bmr024`
   - $K = 9$ participants
   - ≈ 55 minutes
2. train the Θ model
3. compute local perplexity using 60-second Hamming windows



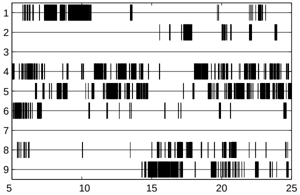
An Example of the Perplexity Trajectory in Time

1. obtain $\mathbf{Q}$ for ICSI `Bmr024`
   - $K = 9$ participants
   - $\approx$ 55 minutes
2. train the $\mathbf{\Theta}$ model
3. compute local perplexity using 60-second Hamming windows



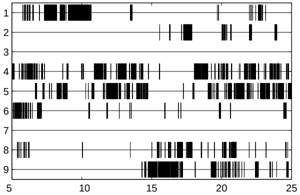
An Example of the Perplexity Trajectory in Time

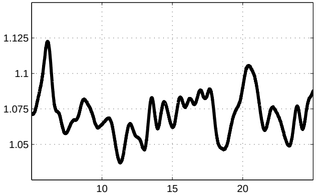
1. obtain **Q** for ICSI `Bmr024`
   - $K = 9$ participants
   - $\approx$ 55 minutes
2. train the $\boldsymbol{\Theta}$ model
3. compute local perplexity using 60-second Hamming windows



An Example of the Perplexity Trajectory in Time

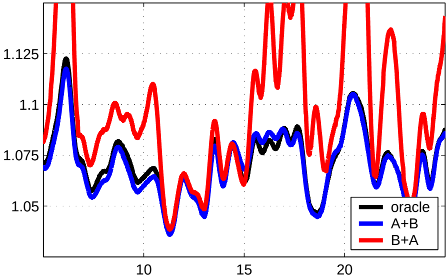
1. obtain $\mathbf{Q}$ for ICSI `Bmr024`
   - $K = 9$ participants
   - ≈ 55 minutes
2. train the $\boldsymbol{\Theta}$ model
3. compute local perplexity using 60-second Hamming windows





Generalization

- **A+B**: train on first half (A) only, test on A
- **B+A**: train on second half (B) only, test on A



- a multi-participant compositional model **Θ** generalizes poorly
- even to other parts of the **same conversation!**

Circumscribing Model Complexity

$$P\left(\mathbf{Q}\right) \;\approx\; P_0 \cdot \prod_{t=1}^{T} P\left(\mathbf{q}_t \,|\, \mathbf{q}_{t-1}, \boldsymbol{\Theta}^{CD}\right) \qquad 2^K \cdot \left(2^K - 1\right)$$

Circumscribing Model Complexity

$$P(\mathbf{Q}) \approx P_0 \cdot \prod_{t=1}^{T} P\left(\mathbf{q}_t \,|\, \mathbf{q}_{t-1}, \boldsymbol{\Theta}^{CD}\right) \qquad 2^K \cdot \left(2^K - 1\right)$$

$$\approx P_0 \cdot \prod_{t=1}^{T} \prod_{k=1}^{K} P\left(\mathbf{q}_t[k] \,|\, \mathbf{q}_{t-1}, \boldsymbol{\Theta}_k^{CI}\right) \qquad K \cdot 2^K$$

Circumscribing Model Complexity

$$
\begin{aligned}
P(\mathbf{Q}) &\approx P_0 \cdot \prod_{t=1}^{T} P\left(\mathbf{q}_t \,|\, \mathbf{q}_{t-1}, \mathbf{\Theta}^{CD}\right) && 2^K \cdot \left(2^K - 1\right) \\
&\approx P_0 \cdot \prod_{t=1}^{T} \prod_{k=1}^{K} P\left(\mathbf{q}_t[k] \,|\, \mathbf{q}_{t-1}, \mathbf{\Theta}_k^{CI}\right) && K \cdot 2^K \\
&\approx P_0 \cdot \prod_{t=1}^{T} \prod_{k=1}^{K} P\left(\mathbf{q}_t[k] \,|\, \mathbf{q}_{t-1}, \mathbf{\Theta}_{any}^{CI}\right) && 2^K
\end{aligned}
$$

Circumscribing Model Complexity

$$P(\mathbf{Q}) \approx P_0 \cdot \prod_{t=1}^{T} P\left(\mathbf{q}_t \mid \mathbf{q}_{t-1}, \mathbf{\Theta}^{CD}\right) \qquad 2^K \cdot \left(2^K - 1\right)$$

$$\approx P_0 \cdot \prod_{t=1}^{T} \prod_{k=1}^{K} P\left(\mathbf{q}_t[k] \mid \mathbf{q}_{t-1}, \mathbf{\Theta}_k^{CI}\right) \qquad K \cdot 2^K$$

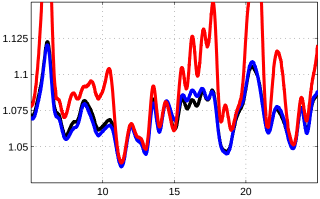
$$\approx P_0 \cdot \prod_{t=1}^{T} \prod_{k=1}^{K} P\left(\mathbf{q}_t[k] \mid \mathbf{q}_{t-1}, \mathbf{\Theta}_{any}^{CI}\right) \qquad 2^K$$

$$\approx P_0 \cdot \prod_{t=1}^{T} \prod_{k=1}^{K} P\left(\mathbf{q}_t[k] \mid \mathbf{q}_{t-1}[k], \mathbf{\Theta}_k^{MI}\right) \qquad K \cdot 2$$

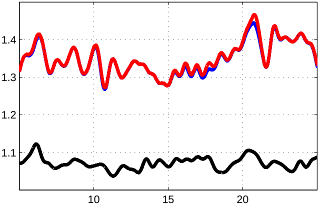
Circumscribing Model Complexity

$$
\begin{aligned}
P(\mathbf{Q}) &\approx P_0 \cdot \prod_{t=1}^{T} P\left(\mathbf{q}_t \,|\, \mathbf{q}_{t-1}, \mathbf{\Theta}^{CD}\right) && 2^K \cdot \left(2^K - 1\right) \\
&\approx P_0 \cdot \prod_{t=1}^{T} \prod_{k=1}^{K} P\left(\mathbf{q}_t\,[k] \,|\, \mathbf{q}_{t-1}, \mathbf{\Theta}_k^{CI}\right) && K \cdot 2^K \\
&\approx P_0 \cdot \prod_{t=1}^{T} \prod_{k=1}^{K} P\left(\mathbf{q}_t\,[k] \,|\, \mathbf{q}_{t-1}, \mathbf{\Theta}_{any}^{CI}\right) && 2^K \\
&\approx P_0 \cdot \prod_{t=1}^{T} \prod_{k=1}^{K} P\left(\mathbf{q}_t\,[k] \,|\, \mathbf{q}_{t-1}\,[k], \mathbf{\Theta}_k^{MI}\right) && K \cdot 2 \\
&\approx P_0 \cdot \prod_{t=1}^{T} \prod_{k=1}^{K} P\left(\mathbf{q}_t\,[k] \,|\, \mathbf{q}_{t-1}\,[k], \mathbf{\Theta}_{any}^{MI}\right) && 2
\end{aligned}
$$

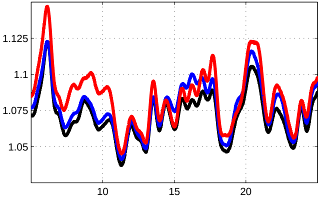
Circumscribing Model Complexity: Perplexity Trajectories



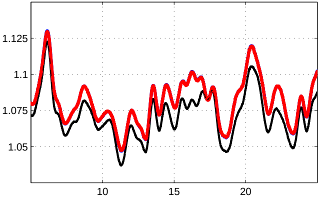
Θ_k^{CI}



Θ_{any}^{CI}



Θ_k^{MI}



Θ_{any}^{MI}

Directly Estimated Compositional Model Limitations

- Direct compositional models either:
  1. Periodically underperform — grossly — due to overfitting, or
  2. Do not model interaction ($\boldsymbol{\Theta}_k^{MI}$, $\boldsymbol{\Theta}_{any}^{MI}$).

- Variants which **do** model interaction ($\boldsymbol{\Theta}^{CD}$, $\boldsymbol{\Theta}_k^{CI}$, $\boldsymbol{\Theta}_{any}^{CI}$):
  1. Fail to exhibit *K-independence*:
     - the number and identity of states is a function of K
     - cannot be trained on conversations with K participants, and applied to conversations with $K' \neq K$ participants
  2. Fail to exhibit *R-independence*:
     - sensitive to participant index assignment
     - perplexities differ if Q is rotated by arbitrary rotation R
     - exhaustive rotation during training has complexity $K!$
  3. Insufficiently parsimonious — theoretically vacuous

Directly Estimated Compositional Model Limitations

- Direct compositional models either:
  1. Periodically underperform — grossly — due to overfitting, or
  2. Do not model interaction ($\mathbf{\Theta}_k^{MI}$, $\mathbf{\Theta}_{any}^{MI}$).
- Variants which **do** model interaction ($\mathbf{\Theta}^{CD}$, $\mathbf{\Theta}_k^{CI}$, $\mathbf{\Theta}_{any}^{CI}$):
  1. Fail to exhibit *K-independence*.
     - the number and identity of states is a function of K
     - cannot be trained on conversations with K participants, and applied to conversations with $K' \neq K$ participants
  2. Fail to exhibit $\mathbf{R}$-*independence*.
     - sensitive to participant index assignment
     - perplexities differ if $\mathbf{Q}$ is rotated by arbitrary rotation $\mathbf{R}$
     - exhaustive rotation during training has complexity $K!$
  3. Insufficiently parsimonious $\longrightarrow$ theoretically vacuous.

Directly Estimated Compositional Model Limitations

- Direct compositional models either:
  1. Periodically underperform — grossly — due to overfitting, or
  2. Do not model interaction ($\mathbf{\Theta}_k^{MI}$, $\mathbf{\Theta}_{any}^{MI}$).

- Variants which **do** model interaction ($\mathbf{\Theta}^{CD}$, $\mathbf{\Theta}_k^{CI}$, $\mathbf{\Theta}_{any}^{CI}$):
  1. Fail to exhibit K-*independence*.
     - the number and identity of states is a function of K
     - cannot be trained on conversations with K participants, and applied to conversations with $K' \neq K$ participants
  2. Fail to exhibit **R**-*independence*.
     - sensitive to participant index assignment
     - perplexities differ if $\mathbf{Q}$ is rotated by arbitrary rotation $\mathbf{R}$
     - exhaustive rotation during training has complexity $K!$
  3. Insufficiently parsimonious $\longrightarrow$ theoretically vacuous.

Directly Estimated Compositional Model Limitations

- Direct compositional models either:
  1. Periodically underperform — grossly — due to overfitting, or
  2. Do not model interaction ($\mathbf{\Theta}_k^{MI}$, $\mathbf{\Theta}_{any}^{MI}$).
- Variants which **do** model interaction ($\mathbf{\Theta}^{CD}$, $\mathbf{\Theta}_k^{CI}$, $\mathbf{\Theta}_{any}^{CI}$):
  1. Fail to exhibit *K-independence*.
     - the number and identity of states is a function of K
     - cannot be trained on conversations with K participants, and applied to conversations with $K' \neq K$ participants
  2. Fail to exhibit **R**-*independence*.
     - sensitive to participant index assignment
     - perplexities differ if **Q** is rotated by arbitrary rotation **R**
     - exhaustive rotation during training has complexity $K!$
  3. Insufficiently parsimonious $\longrightarrow$ theoretically vacuous.

Degree-of-Overlap (DO) Model

Replace the probability of transition between **compositional states** by the probability of transition between **the number of participants speaking simultaneously** in them:

$$P\left(\mathbf{q}_t \,|\, \mathbf{q}_{t-1}, \boldsymbol{\Theta}^{CD}\right) \doteq \alpha\, P\left(\|\mathbf{q}_t\| \,|\, \|\mathbf{q}_{t-1}\|, \boldsymbol{\Theta}^{DO}\right)$$

where

$$\begin{aligned} \|\mathbf{q}\| &= \sum_{k=1}^{K} \delta\left(\mathbf{q}[k], \blacksquare\right) \\ &\in \{0, 1, \cdots, K\} \end{aligned}$$

Model contains only $K \cdot (K+1)$ free parameters.

DO Model Examples

$$\begin{bmatrix} \square \\ \square \\ \square \end{bmatrix} \rightarrow \begin{bmatrix} \square \\ \square \\ \blacksquare \end{bmatrix} \qquad \begin{bmatrix} \blacksquare \\ \blacksquare \\ \square \end{bmatrix} \rightarrow \begin{bmatrix} \square \\ \blacksquare \\ \square \end{bmatrix}$$

$$0 \rightarrow 1 \qquad\qquad 2 \rightarrow 1$$

But unfortunately,

$$\begin{bmatrix} \square \\ \blacksquare \\ \square \end{bmatrix} \rightarrow \begin{bmatrix} \square \\ \blacksquare \\ \square \end{bmatrix} \qquad \begin{bmatrix} \square \\ \blacksquare \\ \square \end{bmatrix} \rightarrow \begin{bmatrix} \blacksquare \\ \square \\ \square \end{bmatrix}$$

$$1 \rightarrow 1 \qquad\qquad 1 \rightarrow 1$$

Extended-Degree-of-Overlap Model (Laskowski & Schultz, 2007)

Extend the "to" state,

- to a 2-element vector, with the
- **number of participants speaking in both $\mathbf{q}_{t-1}$ and $\mathbf{q}_t$:**

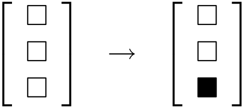
$$P\left(\mathbf{q}_t \,|\, \mathbf{q}_{t-1}, \mathbf{\Theta}^{CD}\right) \doteq \alpha\, P\left(\|\mathbf{q}_t\|, \|\mathbf{q}_t \cdot \mathbf{q}_{t-1}\| \,|\, \|\mathbf{q}_{t-1}\|, \mathbf{\Theta}^{EDO}\right)$$

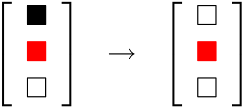
where

$$\left(\mathbf{q} \cdot \mathbf{q}'\right)[k] = \begin{cases} \blacksquare & \text{if } \mathbf{q}[k] = \blacksquare \text{ and } \mathbf{q}'[k] = \blacksquare \\ \square & \text{otherwise} \end{cases}$$

Also easy to train: just count the bigrams!

EDO Model Examples

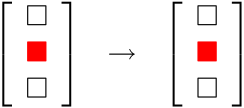


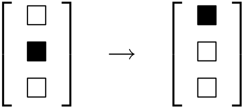


$0 \rightarrow [\,0, 1\,]$

$2 \rightarrow [\,1, 1\,]$

And as desired,





$1 \rightarrow [\,1, 1\,]$

$1 \rightarrow [\,0, 1\,]$

EDO Desiderata Scorecard

1. The EDO Model achieves **R**-invariance:
   - $\|\cdot\|$ is a sum
   - commutative $\longrightarrow$ rotation-independent
   - **results same regardless of participant index assignment**

2. The EDO Model achieves K-invariance:
   - sums performed over ■-state participants only
   - remaining participants, in □, ignored
   - **can apply to any K, with $K_{train} \neq K_{test}$**

3. The EDO state space is tractably small:
   - parameters are individually meaninful
   - can be further constrained by mapping all $\|\cdot\| > K_{max}$ to K_{max}

EDO Desiderata Scorecard

1. The EDO Model achieves **R**-invariance:
   - $\|\cdot\|$ is a sum
   - commutative $\longrightarrow$ rotation-independent
   - **results same regardless of participant index assignment**
2. The EDO Model achieves K-invariance:
   - sums performed over ■-state participants only
   - remaining participants, in □, ignored
   - **can apply to any** K**, with** $K_{train} \neq K_{test}$
3. The EDO state space is tractably small:
   - parameters are individually meaninful
   - can be further constrained by mapping all $\|\cdot\| > K_{max}$ to K_{max}

EDO Desiderata Scorecard

1. The EDO Model achieves **R**-invariance:
   - $\| \cdot \|$ is a sum
   - commutative $\longrightarrow$ rotation-independent
   - **results same regardless of participant index assignment**
2. The EDO Model achieves K-invariance:
   - sums performed over $\blacksquare$-state participants only
   - remaining participants, in $\square$, ignored
   - **can apply to any** K**, with** $K_{train} \neq K_{test}$
3. The EDO state space is tractably small:
   - parameters are individually meaninful
   - can be further constrained by mapping all $\| \cdot \| > K_{max}$ to K_{max}

Data

ICSI Meeting Corpus (Janin et al, 2003; Shriberg et al, 2004):

- 75 meetings
- would have occurred even if they had not been recorded
- approximately 1 hour long
- 3–9 participants each
- forced-alignment-mediated ■/□ references available

Same-Conversation Training

- iterate over all meetings:
  1. split meeting into halves A and B
  2. A+B condition: { train A, score A } & { train B, score B }
  3. B+A condition: { train A, score B } & { train B, score A }
- scoring intervals of the **same** conversation
  - number of participants K invariable
  - participant index assignment $\mathbf{R}$ invariable
- assess
  - independent-participant model $\mathbf{\Theta}^{MI}_{any}$
  - compositional models: $\mathbf{\Theta}^{CD}$, $\mathbf{\Theta}^{CI}_{k}$, $\mathbf{\Theta}^{MI}_{k}$
  - EDO model, with $K_{max} = K$

Same-Conversation Results

Model	PP, A+B		PP, B+A	
	"all"	"sub"	"all"	"sub"
oracle	1.0905	1.6444	1.0905	1.6444
Θ^{CD}	**1.0905**	**1.6444**	1.1225	1.8395
$\{\Theta_k^{CI}\}$	1.0915	1.6576	1.1156	1.7809
$\{\Theta_k^{MI}\}$	1.0978	1.7236	1.1086	1.7950
Θ^{MI}	1.1046	1.8047	1.1047	1.8059
Θ^{EDO}	1.0977	1.7257	**1.0985**	**1.7323**

- on unseen same-conversation data, EDO model outperforms all compositional models

Same-Conversation Results

Model	PP, A+B		PP, B+A	
	"all"	"sub"	"all"	"sub"
oracle	1.0905	1.6444	1.0905	1.6444
$\mathbf{\Theta}^{CD}$	**1.0905**	**1.6444**	1.1225	1.8395
$\{\mathbf{\Theta}_k^{CI}\}$	1.0915	1.6576	1.1156	1.7809
$\{\mathbf{\Theta}_k^{MI}\}$	1.0978	1.7236	1.1086	1.7950
$\mathbf{\Theta}^{MI}$	1.1046	1.8047	1.1047	1.8059
$\mathbf{\Theta}^{EDO}$	1.0977	1.7257	**1.0985**	**1.7323**

- on unseen same-conversation data, EDO model outperforms all compositional models

Other-Conversation Training

- iterate over all meetings:
  1. train on remaining 74 meetings
  2. score held out meeting

- scoring **different** conversations
  - number of participants K **variable**
  - participant index assignment **R unknown**

- assess
  - independent-participant model $\mathbf{\Theta}^{MI}_{any}$
  - EDO model, over a range of K_{max}

Other-Conversation Results

Model	PP		ΔPP (%)	
	"all"	"sub"	"all"	"sub"
oracle	1.0921	1.6616	−100	−100
$\mathbf{\Theta}^{MI}$	1.1051	1.8170	0	0
$\mathbf{\Theta}^{EDO}$ (6)	1.0992	1.7405	−45	−49
$\mathbf{\Theta}^{EDO}$ (5)	1.0968	1.7127	−64	−67
$\mathbf{\Theta}^{EDO}$ (4)	**1.0953**	**1.6947**	−75	−79
$\mathbf{\Theta}^{EDO}$ (3)	1.1082	1.8502	+24	+21

Conclusions

1. **The Extended-Degree-of-Overlap (EDO) model**:
   - can be used as a density estimator for *any* conversation;
   - *any* conversation can be used to infer its parameters.
2. **The EDO model vastly outperforms standard single-participant alternatives**,
   - e.g., those used in speech activity detection,
   - by 75%rel from the oracle.
3. **Participant behavior is** (in measurable part) **predicted by interlocutor behavior**.

Conclusions

1. **The Extended-Degree-of-Overlap (EDO) model**:
   - can be used as a density estimator for *any* conversation;
   - *any* conversation can be used to infer its parameters.
2. **The EDO model vastly outperforms standard single-participant alternatives**,
   - e.g., those used in speech activity detection,
   - by 75%rel from the oracle.
3. **Participant behavior is** (in measurable part) **predicted by interlocutor behavior**.

Conclusions

1. **The Extended-Degree-of-Overlap (EDO) model**:
   - can be used as a density estimator for *any* conversation;
   - *any* conversation can be used to infer its parameters.
2. **The EDO model vastly outperforms standard single-participant alternatives**,
   - e.g., those used in speech activity detection,
   - by 75%rel from the oracle.
3. **Participant behavior is** (in measurable part) **predicted by interlocutor behavior**.

Contributions

1. A **framework for computing perplexity** in ■/□ interaction chronographs;
2. Evidence of the **unsuitability of directly estimated compositional models**;
3. Evidence of the **suitability of the EDO model** as a baseline for future research; and
4. Empirical assessment of EDO performance on one of the **largest multi-party corpora of naturally occurring conversation**.

Contributions

1. A **framework for computing perplexity** in ■/□ interaction chronographs;
2. Evidence of the **unsuitability of directly estimated compositional models**;
3. Evidence of the **suitability of the EDO model** as a baseline for future research; and
4. Empirical assessment of EDO performance on one of the **largest multi-party corpora of naturally occurring conversation**.

Contributions

1. A **framework for computing perplexity** in ■/□ interaction chronographs;
2. Evidence of the **unsuitability of directly estimated compositional models**;
3. Evidence of the **suitability of the EDO model** as a baseline for future research; and
4. Empirical assessment of EDO performance on one of the **largest multi-party corpora of naturally occurring conversation**.

Contributions

1. A **framework for computing perplexity** in ■/□ interaction chronographs;
2. Evidence of the **unsuitability of directly estimated compositional models**;
3. Evidence of the **suitability of the EDO model** as a baseline for future research; and
4. Empirical assessment of EDO performance on one of the **largest multi-party corpora of naturally occurring conversation**.

Impact/Recommendations

1. The precise EDO model formulation complements and possibly supersedes the (heretofore usefully) imprecise notion of taking turns.
   - both account for the distribution of speech in time and across participants
2. The EDO model and PPL measure provide a computational means for corroborating the findings of conversation analysis (in particular) on vastly larger collections of conversation than have been analyzed to date.
3. The EDO model and PPL measure provide an unambiguous measure of spoken document similarity; can now easily:
   - compare turn-taking across conversations,
   - find where turn-taking deviates from norms ("hotspots"), and
   - assess turn-taking appropriateness in dialogue systems.

Impact/Recommendations

1. The precise EDO model formulation complements and possibly supersedes the (heretofore usefully) imprecise notion of taking turns.
   - both account for the distribution of speech in time and across participants
2. The EDO model and PPL measure provide a computational means for corroborating the findings of conversation analysis (in particular) on vastly larger collections of conversation than have been analyzed to date.
3. The EDO model and PPL measure provide an unambiguous measure of spoken document similarity; can now easily:
   - compare turn-taking across conversations,
   - find where turn-taking deviates from norms ("hotspots"), and
   - assess turn-taking appropriateness in dialogue systems.

Impact/Recommendations

1. The precise EDO model formulation complements and possibly supersedes the (heretofore usefully) imprecise notion of taking turns.
   - both account for the distribution of speech in time and across participants
2. The EDO model and PPL measure provide a computational means for corroborating the findings of conversation analysis (in particular) on vastly larger collections of conversation than have been analyzed to date.
3. The EDO model and PPL measure provide an unambiguous measure of spoken document similarity; can now easily:
   - compare turn-taking across conversations,
   - find where turn-taking deviates from norms ("hotspots"), and
   - assess turn-taking appropriateness in dialogue systems.

Thank You!