PRELIMINARIES TO AN ACCOUNT OF MULTI-PARTY CONVERSATIONAL TURN-TAKING AS AN ANTIFERROMAGNETIC SPIN GLASS

Kornel Laskowski, Mattias Heldner & Jens Edlund
KTH Speech Music and Hearing, Stockholm, Sweden

Goal

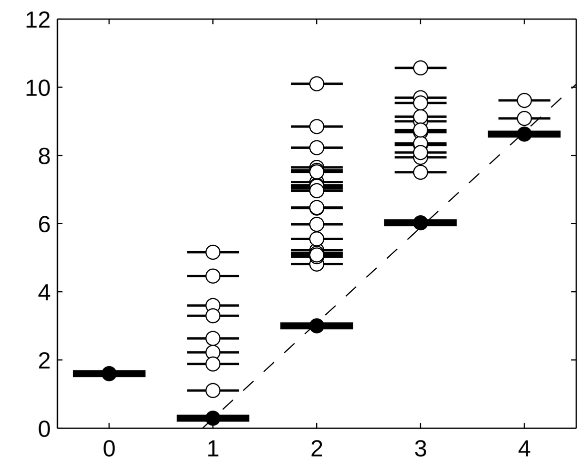
Argue for **a parametric form** which reduces the complexity of a model $\boldsymbol{\Theta}$ of **multi-party turn-taking**.

The role of $\boldsymbol{\Theta}$ is to provide the likelihood of an arbitrary-size **speech/non-speech chronogram**

$$\mathbf{Q} = \left[\begin{array}{c} \square\square\square\square\square\square\blacksquare\blacksquare\square\blacksquare\blacksquare \\ \blacksquare\square\blacksquare\blacksquare\blacksquare\square\square\square\square\square\square \\ \cdots\square\square\square\square\square\square\square\square\square\square\square\cdots \\ \square\square\square\square\blacksquare\blacksquare\square\square\square\square\square \\ \square\square\square\square\square\square\square\square\square\square\square \end{array}\right]$$

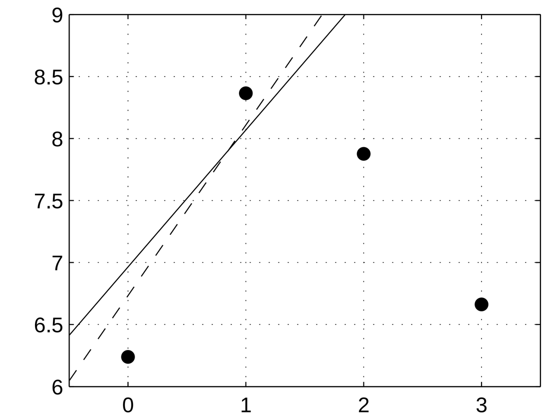
Findings I

1. The log-likelihood of the unconditional probability of specific degree $\boldsymbol{v}$ of overlap is linear in $\boldsymbol{v}$.

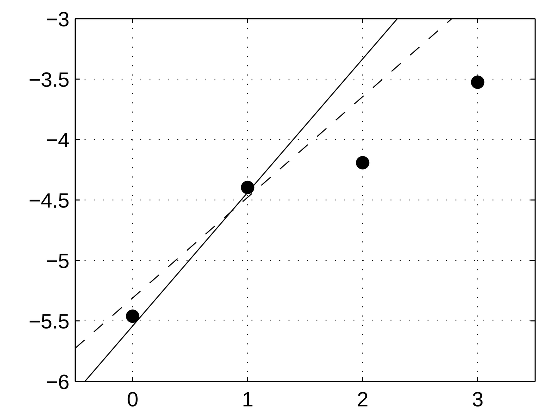


Findings II

2. The logit-likelihood of the conditional probability of speaking given a specific degree $\boldsymbol{v}$ of already-occurring overlap is *nearly* linear in $\boldsymbol{v}$.



entry



egress

Potential Impact

I. Significant reduction in number of parameters:

Conversation-specific form:
$$\boldsymbol{K}\left(\mathbf{2}^{\boldsymbol{mK}}\right) \mapsto \boldsymbol{K} + \boldsymbol{mK}^{\mathbf{2}}$$

Conversation-independent form:
$$\left(\mathbf{2}\boldsymbol{K}\right)^{\boldsymbol{m}} \mapsto \mathbf{1} + \mathbf{2}\boldsymbol{m}$$

II. The parametric model is a neural network:

Predictions can easily be extended to rely on *any* feature type at instant $\boldsymbol{t} - \mathbf{1}$.

III. Model structure is directly interpretable.

Pairwise influence overtly coded in the magnitude of the off-diagonal entries of **W**.

Standard Non-Parametric Form

Factor in time, Markovian assumption of order $\boldsymbol{m} = \mathbf{1}$:

$$\boldsymbol{P}\left(\begin{array}{c} \square\square\square\square\square\square\blacksquare\blacksquare\square\blacksquare\blacksquare \\ \blacksquare\square\blacksquare\blacksquare\blacksquare\square\square\square\square\square\square \\ \cdots\square\square\square\square\square\square\square\square\square\square\square\cdots \\ \square\square\square\square\blacksquare\blacksquare\square\square\square\square\square \\ \square\square\square\square\square\square\square\square\square\square\square \end{array}\right) = \cdots \times \boldsymbol{P}\left(\begin{array}{c|c} \square & \square \\ \blacksquare & \blacksquare \\ \square & \square \\ \blacksquare & \square \\ \square & \square \end{array}\right) \times \boldsymbol{P}\left(\begin{array}{c|c} \square & \square \\ \square & \blacksquare \\ \square & \square \\ \blacksquare & \blacksquare \\ \square & \square \end{array}\right) \times \boldsymbol{P}\left(\begin{array}{c|c} \blacksquare & \square \\ \square & \square \\ \square & \square \\ \square & \blacksquare \\ \square & \square \end{array}\right) \times \cdots$$

Factor each factor, assuming participants are *conditionally independent*:

$$\boldsymbol{P}\left(\begin{array}{c|c} \square & \square \\ \square & \blacksquare \\ \square & \square \\ \blacksquare & \blacksquare \\ \square & \square \end{array}\right) = \boldsymbol{P}_1\left(\begin{array}{c|c} {\scriptstyle 1}\,\square & \square \\ {\scriptstyle 2} & \blacksquare \\ {\scriptstyle 3} & \square \\ {\scriptstyle 4} & \blacksquare \\ {\scriptstyle 5} & \square \end{array}\right) \times \boldsymbol{P}_2\left(\begin{array}{c|c} {\scriptstyle 1} & \square \\ {\scriptstyle 2}\,\square & \blacksquare \\ {\scriptstyle 3} & \square \\ {\scriptstyle 4} & \blacksquare \\ {\scriptstyle 5} & \square \end{array}\right) \times \boldsymbol{P}_3\left(\begin{array}{c|c} {\scriptstyle 1} & \square \\ {\scriptstyle 2} & \blacksquare \\ {\scriptstyle 3}\,\square & \square \\ {\scriptstyle 4} & \blacksquare \\ {\scriptstyle 5} & \square \end{array}\right) \times \boldsymbol{P}_4\left(\begin{array}{c|c} {\scriptstyle 1} & \square \\ {\scriptstyle 2} & \blacksquare \\ {\scriptstyle 3} & \square \\ {\scriptstyle 4}\,\blacksquare & \blacksquare \\ {\scriptstyle 5} & \square \end{array}\right) \times \boldsymbol{P}_5\left(\begin{array}{c|c} {\scriptstyle 1} & \square \\ {\scriptstyle 2} & \blacksquare \\ {\scriptstyle 3} & \square \\ {\scriptstyle 4} & \blacksquare \\ {\scriptstyle 5}\,\square & \square \end{array}\right)$$

This is the **conversation-specific** model. It has $\boldsymbol{K}$ submodels, each of $\mathbf{2} \cdot \mathbf{2}^{\boldsymbol{K}}$ parameters $\longrightarrow \boldsymbol{K} \cdot \left(\mathbf{2}^{\boldsymbol{K}}\right)^{\boldsymbol{m}}$ degrees of freedom in total.

For the **conversation-independent** model,

$$\boldsymbol{P}\left(\begin{array}{c|c} \square & \square \\ \square & \blacksquare \\ \square & \square \\ \blacksquare & \blacksquare \\ \square & \square \end{array}\right) = \boldsymbol{P}\left(\begin{array}{c|c} {\scriptstyle 1}\,\square & \square \\ {\scriptstyle 2} & \blacksquare \\ {\scriptstyle 3} & \square \\ {\scriptstyle 4} & \blacksquare \\ {\scriptstyle 5} & \square \end{array}\right) \times \boldsymbol{P}\left(\begin{array}{c|c} {\scriptstyle 2}\,\square & \blacksquare \\ {\scriptstyle 1} & \square \\ {\scriptstyle 3} & \square \\ {\scriptstyle 4} & \blacksquare \\ {\scriptstyle 5} & \square \end{array}\right) \times \boldsymbol{P}\left(\begin{array}{c|c} {\scriptstyle 3}\,\square & \square \\ {\scriptstyle 1} & \square \\ {\scriptstyle 2} & \blacksquare \\ {\scriptstyle 4} & \blacksquare \\ {\scriptstyle 5} & \square \end{array}\right) \times \boldsymbol{P}\left(\begin{array}{c|c} {\scriptstyle 4}\,\blacksquare & \blacksquare \\ {\scriptstyle 1} & \square \\ {\scriptstyle 2} & \blacksquare \\ {\scriptstyle 3} & \square \\ {\scriptstyle 5} & \square \end{array}\right) \times \boldsymbol{P}\left(\begin{array}{c|c} {\scriptstyle 5}\,\square & \square \\ {\scriptstyle 1} & \square \\ {\scriptstyle 2} & \blacksquare \\ {\scriptstyle 3} & \square \\ {\scriptstyle 4} & \blacksquare \end{array}\right)$$

$$= \boldsymbol{P}\left(\begin{array}{c|c} {\scriptstyle 1}\,\square & \square \\ {\scriptstyle oth} & \mathbf{2} \end{array}\right) \times \boldsymbol{P}\left(\begin{array}{c|c} {\scriptstyle 2}\,\square & \blacksquare \\ {\scriptstyle oth} & \mathbf{1} \end{array}\right) \times \boldsymbol{P}\left(\begin{array}{c|c} {\scriptstyle 3}\,\square & \square \\ {\scriptstyle oth} & \mathbf{2} \end{array}\right) \times \boldsymbol{P}\left(\begin{array}{c|c} {\scriptstyle 4}\,\blacksquare & \blacksquare \\ {\scriptstyle oth} & \mathbf{1} \end{array}\right) \times \boldsymbol{P}\left(\begin{array}{c|c} {\scriptstyle 5}\,\square & \square \\ {\scriptstyle oth} & \mathbf{2} \end{array}\right)$$

There is only 1 submodel, of $\mathbf{2} \cdot (\mathbf{2} \cdot \boldsymbol{K})$ parameters $\longrightarrow (\mathbf{2} \cdot \boldsymbol{K})^{\boldsymbol{m}}$ degrees of freedom in total.

A "Natural" Consequence of the Logistic Fit

Logistic regression $\equiv$ Glauber dynamics $\equiv$ most prosaic feed-forward neural network structure:

$$\boldsymbol{P}_k\left(\mathbf{q}_t[k] = \blacksquare \,|\, \mathbf{q}_{t-1}\right) = \frac{\mathbf{1}}{\mathbf{1} + \boldsymbol{e}^{-\beta \mathbf{h}_{t-1}[k]}}$$

$$\boldsymbol{P}_k\left(\mathbf{q}_t[k] = \square \,|\, \mathbf{q}_{t-1}\right) = \mathbf{1} - \boldsymbol{P}_k\left(\mathbf{q}_t[k] = \blacksquare \,|\, \mathbf{q}_{t-1}\right)$$

$$\mathbf{h}_{t-1} = \mathbf{b} + \mathbf{W} \cdot \mathbf{q}_{t-1} \quad \text{with} \quad \{\square, \blacksquare\} \equiv \{\mathbf{0}, \mathbf{1}\}$$

Conversation-specific model:

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{bmatrix} \qquad \mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & w_{13} & w_{14} & w_{15} \\ w_{21} & w_{22} & w_{23} & w_{24} & w_{25} \\ w_{31} & w_{32} & w_{33} & w_{34} & w_{35} \\ w_{41} & w_{42} & w_{43} & w_{44} & w_{45} \\ w_{51} & w_{52} & w_{53} & w_{54} & w_{55} \end{bmatrix}$$

$\boldsymbol{K} + \boldsymbol{mK}^{\mathbf{2}}$ parameters in total (and degrees of freedom)

This happens to be an ***infinite-range anti-ferromagnetic spin glass (with slow connections)***.

Conversation-independent model:

$$\mathbf{b} = \begin{bmatrix} b \\ b \\ b \\ b \\ b \end{bmatrix} \qquad \mathbf{W} = \begin{bmatrix} w_+ & w_- & w_- & w_- & w_- \\ w_- & w_+ & w_- & w_- & w_- \\ w_- & w_- & w_+ & w_- & w_- \\ w_- & w_- & w_- & w_+ & w_- \\ w_- & w_- & w_- & w_- & w_+ \end{bmatrix}$$

$\mathbf{1} + \mathbf{2}\boldsymbol{m}$ parameters in total (and degrees of freedom)

This happens to be an ***infinite-range Ising anti-ferromagnet with slow connections***.

kornel@cs.cmu.edu, mattias@speech.kth.se, edlund@speech.kth.se