ON THE CONVERSANT-SPECIFICITY OF STOCHASTIC TURN-TAKING MODELS

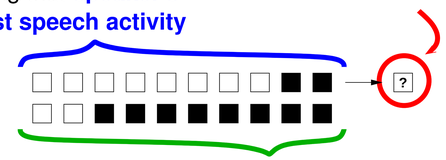
Kornel Laskowski
Carnegie Mellon University, Pittsburgh PA, USA
Voci Technologies, Inc., Pittsburgh PA, USA

Goals

STOCHASTIC TURN-TAKING MODELS

predict a speaker's **future speech activity**

using that **speaker's past speech activity**



and their **interlocutor's past speech activity**

QUESTIONS:

1. Is variability in trained models accounted for by speaker identity?
2. **Can we infer speaker identity from trained models?**

Findings

1. Although there is "no data like more data", stochastic turn-taking models are **surprisingly speaker-specific**.
2. Variability between person-specific models is **correlated with person identity**.
3. This variability is large enough to permit use of these models for:
   - **speaker attribution**: better than a task-specific baseline by 45%rel on unseen data
   - **speaker detection**: better than a task-specific baseline by 36%rel on unseen data

Note that task-specific baselines also relied on the **unconditional probability of speaking**, which these stochastic turn-taking models do no.

Conclusions

1. People generally exhibit an **interactional-timing self-similarity**, even in longitudinal collections.
2. People generally exhibit an **interactional-timing dissimilarity to other people**.
3. Stochastic turn-taking models are able to **quantify these similarities and dissimilarities**.
4. Stochastic turn-taking models quantify aspects which appear to be independent of the unconditional probability of speaking, or **talkativeness**.

Potential Impact

I. Stochastic turn-taking models appear to measure aspects of **conversational social personality**, based on a very simple representation of conversations.

II. They permit future study of **behavior across different accents, languages, and topics**, with no additional technical overhead.

III. They may be useful in the design of spoken dialogue system agents which exhibit a consistent personality.

IV. They may be suitable for speaker recognition applications in **degraded-signal and privacy-sensitive environments**.

Two Diagnostic Tasks

- **Task 1: Speaker Attribution**
  - know the K identities of the K persons in a conversation
  - do not know which identity goes with each speech stream
  - random guessing: **80%** classification error
- **Task 2: Speaker Detection**
  - do not know the identities of the K persons in a conversation
  - must draw from a larger population of $N > K$ persons
  - random guessing: **93%** classification error
- **Data**
  - 67 meetings from the ICSI Meeting Corpus
  - split into TRAINSET, DEVSET, and (unseen) TESTSET
  - 14 persons occur frequently; all others pooled into UNK

Task-Specific Baseline

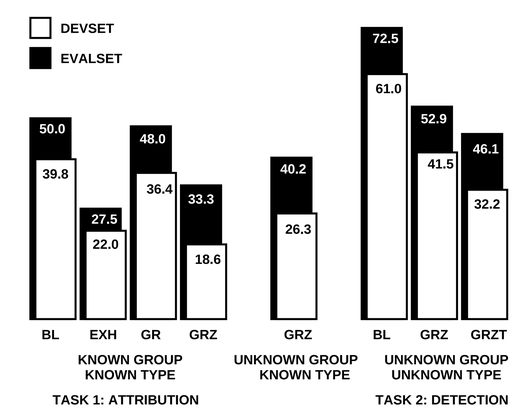
- **Features** (per participant)
  - measure the unconditional probability of speaking
  - measure the conditional probability of speaking
    - conditioning context consists of one 100-ms frame back in time
    - conditioning context models a conversant's behavior wrt to a specific other conversant's behavior
- **Models**
  - one-Gaussian models of probability-of-speaking
  - Group Behavior model
    - product of individual hypothesized conversants probability-of-speaking
  - Group Membership model
    - penalize hypothesized conversants which appear infrequently in the data
- **Search**
  - exhaustive search of $K!$ group assignments for Speaker Attribution
  - greedy search for Speaker Detection

Evaluation of STT Models

- Given a K-row chronogram representing the behavior of K conversants:
  - form K chronograms with two rows
  - row one: specific to conversant k
  - row two: logical-OR of interlocutors of conversant k
- Train an N-gram model for each of N identities in TRAINSET
- over ten 100-ms frames from the modeled speaker and ten 100-ms frames from the interlocutor row $\rightarrow$ a 21-gram
- For a test conversation, apply each of N models to each of K test conversants
- Greedy search over $N \times K$ matrix

Numerical Results

Attribution/detection error in %:



- "BL": baseline (2008)
- "GRZ"/"GRZT": final results of current work

kornel@cs.cmu.edu