A Deep Learning Approach to Modeling Empathy in Addiction Counseling

James Gibson[1], Dogan Can[1], Bo Xiao[1], Zac E. Imel[2], David C. Atkins[3], Panayiotis Georgiou[1], Shrikanth Narayanan[1]
[1]Signal Analysis and Interpretation Lab, University of Southern California, Los Angeles, CA, USA
[2]Department Educational Psychology, University of Utah, Salt Lake City, UT, USA
[3]Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, USA
[1]sail.usc.edu, [2]zac.imel@utah.edu, [3]datkins@u.washington.edu

This work was supported by NSF and NIH

Introduction

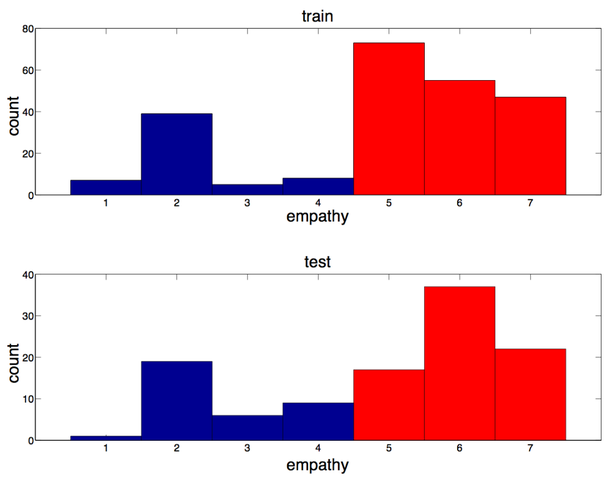
- Motivation
  - Scaling up psychotherapy evaluation
  - Empathy is an important measure for therapist efficacy
- Major questions:
  - How to relate therapist-patient interactions with gestalt behavioral constructs?
  - How to relate speaker turns to local (turn) level behaviors and global (session) level behaviors?
  - How to jointly model these complex phenomena?

Data

- 345 client/counselor interactions (motivational interviews)
- Training and testing sets (~70:30 ratio)
- Session level (MITI) and utterance level (MISC) behavioral codes
- Empathy score (MITI) distribution: 1-7 Likert scale
- MISC28
  - 28 Utterance level behavioral acts
    - 19 counselor, 9 client
    - e.g., reflections, giving information, questions, etc.
- MISC8
  - 8 most frequent codes: 7 counselor, 1 client

Methodology

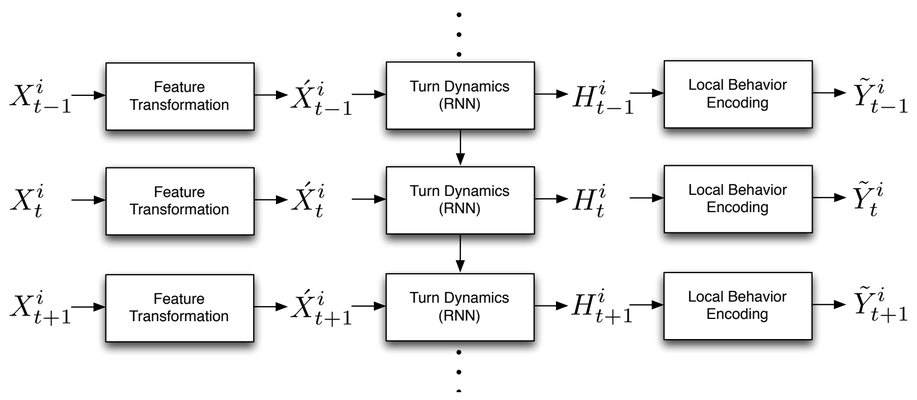
- High vs. low empathy classification



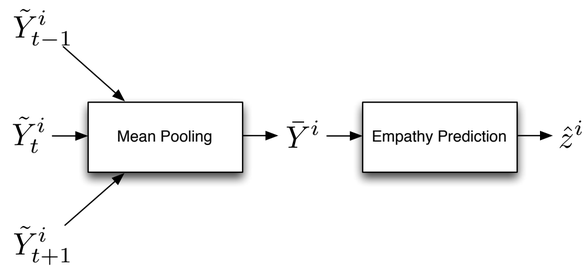
- Multilabel encoding of local behaviors

t=0	Client: I wouldn't mind coming here for treatment but I don't want to go to one of those places where everyone sits around crying and complaining all day.	CLI	[0 0 0 0 0 0 0 1]
t=1	Counselor: You don't want that. So you are kind of wondering what it would be like here.	RES, REC	[0 0 0 0 1 1 0 0]

- Features
  - Word embedding vectors averaged across turn
- Deep learning encoder-decoder network
  - Encoder: maps speaker turns to local behaviors



- Decoder: maps local behaviors to global behavior



Experimental Results

- MISC multilabel prediction
  - MISC8: most frequently occurring behavioral acts
  - MISC28: all labeled behavioral acts

code	recall	precision	F1-score
MISC8	0.617	0.675	0.643
MISC28	0.228	0.348	0.258

- Empathy prediction

model	L	UAR (%)
baseline	N/A	71.8
reference	8	73.6
	28	79.6
proposed system w/o pre-training	8	65.0
	28	62.9
proposed system w pre-training	8	78.6
	28	72.9

Conclusion

- The proposed system outperforms a baseline neural network for predicting counselor empathy ratings
- Two-stage training allows for deeper learning
- Future Work
  - RNN language model
  - ASR derived lexical features
  - Prosodic and spectral features
  - Attention mechanism