Quantitative Observational Practice in Family Studies (1 of 3)

Panayiotis (Panos) Georgiou, Shri Narayanan, Gayla Margolin, Brian Baucum, Matt Black, James Gibson, Nassos Katsamanis, Jeremy Lee, & Bo Xiao

Signal processing for Communication Understanding and Behavior Analysis (SCUBA)
Signal Analysis and Interpretation Laboratory (SAIL)
Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089
georgiou@bpi.usc.edu

**Abstract**

**Goal:**
- Transform observational behavior analysis
- Through computational framework
- Modeling of emotionally-rich human interactions
- Signal processing and machine learning
- Existing family therapy data
- Alleviate the tedium of manual annotation
- Offer new analysis capabilities and empower the mental health experts

**Significance:** USA-10mil people receive psychotherapy every year and state of the art hasn’t changed for decades

---

**Approaches**

+ This poster: [- Other two posters]
  - Model interlocutors independently:
    - Lexical, acoustic and visual modalities
  - Model dynamics of interlocutors
  - Incorporate saliency

**Data**

**Couple Therapy Corpus**
- 117 real distressed couples
- 10-minute dyadic interactions
- 596 sessions (96 hours)

**Audio/Lexical and Visual subsets used**
- Use top/bottom 20% for audio, lexical and 25% for video
- Choose subsets with acceptable audio/video qualities
- Used 6 codes with highest human agreement
- Some distributions skewed and not very separable

**Acoustic Classification**

**Q: Does acoustic channel capture behavior?**
- Frame-level low-level descriptors (LLDs)
  - Prosodic: speech/min-vowel-speech, rate, IO, intensity
  - Spectral: 15 MFCCs, 8 MFDBs
  - Voice quality: jitter, shimmer
- Separate features for (wife, husband, all)
- 7 temporal granularities
  - Global: entire session
  - Halves: 2nd half – 1st half
  - Hierarchical: 0.1s, 0.5s, 1s, 5s, 10s windows
- 14 static functions (e.g., mean, std. dev.)

**Results with logistic regression (L2-regularized)**

**Visual Fusions**

**Q: Does lexical channel capture behavior?**
- Test from reference text
- Test from (unoptimized) ASR output

**Example Transcript**

**Head motion modeling**

**Q: Does head motion capture behavior?**

**Overview of the system flow**

**Method**
- Head motion: face recognition & feature point tracking
- Motion event: moving window of 2 sec long, 1 sec shift
- Motion model: linear prediction coeff. (10 order LSF) & power spectrum (2nd to 16th point in 128-FFT, <3.5 Hz)
- Motion clustering: K-means with K = {4, 5, ..., 25}
- Feature: counts of motion events (kinemes) in each cluster
- Classification: linear support vector classifier

**Case study: M2/Wife/Blame**
- Power spectrum of cluster centroids, test mean diff (ANOVA)
- Red/blue = high/low blame, width = test significance

**Data split**
- Middle 50% of each code — training head motion model
- Upper and bottom 25% of each code — binary classification

**Future Work Highlights**
- Introduce “latent layer” of behavioral primitives
- Improve on individual modalities, e.g., optimize ASR
- Implement fusion based on modality saliences

---

**Acknowledgments**

Full list of publications at http://scuba.usc.edu
Work funded by NSF SHB program

---

**References**

- Panayiotis (Panos) Georgiou, Shri Narayanan, Gayla Margolin, Brian Baucum, Matt Black, James Gibson, Nassos Katsamanis, Jeremy Lee, & Bo Xiao

Signal processing for Communication Understanding and Behavior Analysis (SCUBA)
Signal Analysis and Interpretation Laboratory (SAIL)
Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089
georgiou@bpi.usc.edu

---

**Figures**

- Female
- Male
- Average
- Comparison of average accuracies by PS and LSF

---

**Citations**

Full list of publications at http://scuba.usc.edu
Work funded by NSF SHB program
Quantitative Observational Practice in Family Studies (2 of 3)

Panayiotis (Panos) Georgiou, Shri Narayanan, Gayla Margolin, Brian Baucum, Matt Black, James Gibson, Nasso Kitsamanis, Jeremy Lee, & Bo Xiao

Goal:
- Transform observational behavior analysis
- Through computational framework
- Modeling of emotionally-rich human interactions
- Signal processing and machine learning
- Existing family therapy data
- Alleviate the tedium of manual annotation
- Offer new analysis capabilities and empower the mental health experts

Significance: USA-10mil people receive psychotherapy every year and state of the art hasn’t changed for decades

Approaches
- This poster: [- Other two posters]
  - Model interlocutors independently
  - Model dynamics of interlocutors:
    - Acoustic and visual modalities
    - Incorporate Salience

Couple Therapy Corpus
- 117 real distressed couples
- 10-minute dyadic interactions
- 596 sessions (96 hours)

Data used
- Audio/Lexical and Visual subsets used
  - Use top/bottom 20% for audio, lexical and 25% for video
  - Choose subsets with acceptable audio/video qualities
  - Used 6 codes with highest human agreement
  - Some distributions skewed and not very separable

Visual: Head Motion Similarity Measure
- Based on two bags of events and GMM posteriors
- Compute pair-wise KL divergence of events
- Average of small divergence
  - Similar events are salient
  - Variation of other events no effect
- Dynamic change of similarity

Vocal: Features/LLD’s
- Q: Does PCA channel capture behavior?
  - (Implicit speaking): MFCC & Statistical functional

Representative vocal Parameters (55)
- Pitch (5) $[f_0, \Delta f_0, \Delta \Delta f_0]$
- Intensity (3) $[I, \Delta I, \Delta \Delta I]$
- Speech Rate (1) $[\text{syl}]$
- MFCC (20) $[ \mu c_i(0), \sigma c_i(0), i=1,..,13]$

Vocal Entrainment: Validation
Hypothesis 1: Verification: verifying the proposed signal-derived measures capture psychologically-valid notions of entrainment
- Compare real couple interactions with
  - Artificially sequenced interactions
  - Vocal entrainment on real couples higher ✓

Hypothesis 2: Analysis: analyzing the relationship of the vocal entrainment phenomenon and spouses’ affective states
- Compare positive interactions with negative interactions
- Vocal entrainment on positive couples higher ✓

Hypothesis 3: Application: applying vocal entrainment measures as features in a affective state recognition task
- Entrainment correlated with affective behaviors
- Model temporal dynamics of entrainment
- Dataset: Same as Acoustic and Lexical
- Statistical Framework Factorial Hidden Markov Model
- 62.8% accuracy ✓

Vocal: Unsupervised Computational Framework
- Intuitively, "how do two people sound alike as they interact in a conversation?"
- Similarity between two vocal characteristics spaces
- Directional & Symmetric Similarity Measures
- Kullback-Leibler Divergence (KLD) on normalized variance vector or (weighted) angles between PCA directions

Audio/Lexical and Visual subsets used
- Use top/bottom 20% for audio, lexical and 25% for video
- Choose subsets with acceptable audio/video qualities
- Used 6 codes with highest human agreement
- Some distributions skewed and not very separable
**Quantitative Observational Practice in Family Studies (3 of 3)**

Panayiotis (Panos) Georgiou, Shri Narayanan, Gayla Margolin, Brian Baucum, Matt Black, James Gibson, Nassos Katsamanis, Jeremy Lee, & Bo Xiao

Signal processing for Communication Understanding and Behavior Analysis (SCUBA)

Signal Analysis and Interpretation Laboratory (SAIL)

Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089

georgiou@bpi.usc.edu

---

**Goal:**
- Transform observational behavior analysis
- Through computational framework
- Modeling of emotionally-rich human interactions
- Signal processing and machine learning
- Existing family therapy data
- Alleviate the tedium of manual annotation
- Offer new analysis capabilities and empower the mental health experts

**Significance:** USA-10mil people receive psychotherapy every year and state of the art hasn’t changed for decades

---

**Approaches**

- **This poster:** [† Other two posters]
  - Model interlocutors independently
  - Model dynamics of interlocutors
  - Incorporate Salience:
    - Lexical, acoustic and visual modalities

---

**Saliency**

- Couples’ problem solving discussion are rated on a session level
- It is of interest to identify shorter-term events that influence evaluators’ perceptions of the interaction
- These “salient” instances may help to inform both behavioral scientists
- We use multiple instance learning (MIL) to focus on local events in the couples’ therapy sessions

What are the important bits?

---

**Data**

**Couple Therapy Corpus**
- 117 real distressed couples
- 10-minute dyadic interactions
- 596 sessions (96 hours)

---

**Data used**

Audio/Lexical and Visual subsets used
- Use top/bottom 20% for audio, lexical and 25% for video
- Choose subsets with acceptable audio/video qualities
- Used 6 codes with highest human agreement
- Some distributions skewed and not very separable

---

**Behavioral classification through MIL**

**Multiple Instance Learning:**
- We consider each session a “bag” of “instances”
- Instances are varying-length speaker turns or equal-length windows
- Each instance conveys particular behaviors of interest with varying degrees
- MIL is a method for identifying the “salient instances”, i.e., the local events that most greatly affect the final rating assigned to the session

---

**Summary and Future work**

- Explored saliency in MIL framework
- Explored saliency in multiple modalities
- Explored low-level instance features and deriving high-level session features
- Temporal dynamics of salient events for reactivity
- Explore alternative measures for saliency, such as knowledge inspired signal cues (e.g., laughter, crying)

---

**Citations. Acknowledgments**

Full list of publications at http://scuba.usc.edu

Work funded by NSF SHB program