Behavioral Signal Processing
Promise, Challenges and Opportunities for Behavioral Informatics

Shrikanth (Shri) Narayanan
Signal Analysis and Interpretation Laboratory (SAIL)
http://sail.usc.edu/shri.php
Behavior Analysis is Central:
Social Sciences, Commerce, Education, Health, Security...

SOCIAL SCIENCE SCHOLARSHIP
Learning & training
Daily life activities
Behavioral Informatics
User modeling, customer care
Health & well being
Intel, Security & defense

engineering as an enabler
Is the customer on the telephone upset? (only customer side played)
Educational Game: “Cognitive state” Characterization

- CONFIDENT vs. UNCERTAIN

“Uncertainty” manifests itself through combination of vocal, language, and visual behavioral cues

Multimodality!
Distressed Couple therapy
Characterizing affective dynamics, blame patterns
Autism Spectrum Disorders
Characterizing joint attention; quantifying socio-emotional discourse
Social Media Analytics

USC Realtime Sentiment Analysis

From tweets (politics, movies)

Our unique sentiment model and real-time processing infrastructure allow us to gauge live public sentiments toward the 2012 United States presidential candidates as expressed through Twitter. This live demo is work-in-progress as we continue to refine our sentiment model. More information.

Sentiment

- biden
- obama
- paulryan
- romney

Tweet Volumes

Trending words

Go To Tweets/Annotation Page

Mirrored site (in case of outage)
Behavioral Signal Processing (BSP)

Computational methods that model human behavior signals

• manifested in both overt and covert cues
• processed and used by humans explicitly or implicitly
• facilitate human analysis and decision making

Outcome of BSP: “Behavioral informatics”

Quantifying
HUMAN expressed BEHAVIOR
And
HUMAN “felt sense”

Behavior Coding: Humans in the loop

- Support--than supplant--human (expert) analyses

HUMAN BEHAVIOR OR INTERACTION OF INTEREST
(E.G., CHILD INTERACTING WITH A TEACHER)

Direct Observation

DATA CODING

AVAILABLE DATA (E.G., AUDIO, VIDEO, TEXT, PHYSIOLOGICAL)

HUMAN EVALUATOR

JUDGMENTS
(E.G., WHEN IS THE CHILD UNCERTAIN?)

SIGNAL PROCESSING
(E.G., FEATURE EXTRACTION)

COMPUTATIONAL MODELING
(E.G., MACHINE LEARNING)

BEHAVIORAL INFORMATICS

Feedback
Behavioral signal processing

COMPUTING

OF human action and behavior data

FOR meaningful analysis: timely decision making & intervention (action)

BY collaborative integration of human expertise with automated machine processing

HUMANS
A whole range of enabling technologies

- Voice Activity Detection
- Audio Segmentation
- Alignment
- Transcription
- Keyword spotting
- Prosody Modeling: Intonation, Phrasing, Prominence
- Voice Quality
- Natural Language Processing of Text/Transcripts

- Dialog Act Tagging
- Interaction modeling: Turn taking dynamics, Entrainment
- Speaker/Verification Identification
- Affective Computing
- Nonverbal cue modeling
- Joint speech and visual cue processing
- Speaker State and Trait Characterization
EXPRESSIVE BEHAVIOR: USING ACTORS

USC CreativeIT database
- Multimodal database
- USC Engineering and Theatre
- Dyadic Theatrical Improvisations
- Motion Capture, Video, Audio
  - http://sail.usc.edu/improv/

IEMOCAP: Interactive and emotional motion capture database
- Dyadic interaction
- 5 sessions, 2 actors each
- Emotions elicited in context
- ~12 hours of data
  - http://sail.usc.edu/iemocap/

Freely available
KNOWME NETWORKS: PERSONALIZED BEHAVIOR MEASUREMENTS IN REAL LIFE SETTINGS

END-TO-END ENCRYPTION OF SENSITIVE DATA

CHECK RIGHT TO USE SYSTEM FILTER NOISY UPDATES WEB ENABLE DATA ACCESS

WEB SERVER

SQL WITH ENCRYPTED DATA

DATABASE SERVER

DECRIPTION KEYS

DATA WITH SECURE CONNECTION

Nokia N95

3G
GSM
WI-FI

2 ALIVE HEART RATE MONITOR/ACCELEROMETERS (ECG/ACC)

WEB ENABLE DATA ACCESS

DOCTOR

PHYSICAL ACTIVITY, STRESS, COGNITIVE BEHAVIOR IN REALTIME
Who spoke when, for how long, and about what?

Detected Speech Regions

Speakers in the conversation

Transcription

Hello!

How are you?

I am good.

We should go.

Come on!

Ok!
**What more can we infer beyond words from speech?**

<table>
<thead>
<tr>
<th>Speaker:</th>
<th>spkr1 (Doctor)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender:</td>
<td>Male</td>
</tr>
<tr>
<td>Age:</td>
<td>Adult</td>
</tr>
<tr>
<td>Language:</td>
<td>English</td>
</tr>
<tr>
<td>Prominent words:</td>
<td>So ‘n’ your <strong>chest pains</strong> have been going on just for <strong>two days</strong> is that <strong>right</strong></td>
</tr>
<tr>
<td>Prosodic phrasing:</td>
<td>[So ‘n’ your chest] [pains have been going on just for two days] [is that right]</td>
</tr>
<tr>
<td>Dialog act:</td>
<td>Yes-No Question</td>
</tr>
<tr>
<td>Affect:</td>
<td>Neutral</td>
</tr>
<tr>
<td>Attitude:</td>
<td>Polite</td>
</tr>
</tbody>
</table>

Rich information beyond words
Multimodal Biometrics: “who is that”

- **Multimodal Person Identification**
  - Utilizes multiple biometric traits for person identification (face, voice, fingerprint, iris and talking video)
  - mobile Fusion® and secure logon for personal devices

Partner: 3M Cogent Systems
Automatic Dialog Act (DA) recognition: “intent capture”

- Utterance units are defined at intention level (dialog acts)
- Each utterance is assigned a unique Dialog Act label
- e.g., Discourse Annotation and Markup System of Labeling

*Classify utterances based on pragmatic, semantic, syntactic & prosodic information*

Computing aims to detect these types of dialog acts

<table>
<thead>
<tr>
<th>Utterance</th>
<th>Dialog Act</th>
</tr>
</thead>
<tbody>
<tr>
<td>Me, I'm in the legal department.</td>
<td>Statement-non-opinion</td>
</tr>
<tr>
<td>Uh-huh.</td>
<td>Acknowledge (Backchannel)</td>
</tr>
<tr>
<td>I think it's great</td>
<td>Statement-opinion</td>
</tr>
<tr>
<td>Do you have to have any special training?</td>
<td>Yes-No-Question</td>
</tr>
<tr>
<td>Well, how old are you?</td>
<td>Wh-Question</td>
</tr>
<tr>
<td>That's exactly it.</td>
<td>Agree/Accept</td>
</tr>
</tbody>
</table>
Recognizing Emotions?
Expression versus experience

Description
- Categorical (e.g., happy, sad..), Dimensional (arousal, valence)
- Emotion Profiles
- Dynamic descriptions

Role of context

Tracking Emotion Trends using Facial Expressions, Body Language and Speech
Modeling gestures/speech interrelation

VISUAL and VOCAL features

- Speech
  - Prosodic features: Pitch, energy
  - MFCC coefficients (vocal tract)
- Visual features
  - Head motion
  - Eyebrow
  - Lips
  - Different face regions

Multimodal Emotion Recognition

- From speech
  - Average ~70%
  - Confusion sadness-neutral
  - Confusion happiness-anger
- From facial expression
  - Average ~85%
  - Confusion anger-sadness
  - Confusion neutral-happiness
  - Confusion sadness-neutral
- Multimodal system (feature-level)
  - Average ~90%
  - Confusion neutral-sadness
  - Other pairs are correctly separated

Using SVM

<table>
<thead>
<tr>
<th></th>
<th>Anger</th>
<th>Sadness</th>
<th>Happiness</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>0.68</td>
<td>0.05</td>
<td>0.21</td>
<td>0.05</td>
</tr>
<tr>
<td>Sadness</td>
<td>0.07</td>
<td>0.64</td>
<td>0.06</td>
<td>0.22</td>
</tr>
<tr>
<td>Happiness</td>
<td>0.19</td>
<td>0.04</td>
<td>0.70</td>
<td>0.08</td>
</tr>
<tr>
<td>Neutral</td>
<td>0.04</td>
<td>0.14</td>
<td>0.01</td>
<td>0.81</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Anger</th>
<th>Sadness</th>
<th>Happiness</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>0.79</td>
<td>0.18</td>
<td>0.00</td>
<td>0.03</td>
</tr>
<tr>
<td>Sadness</td>
<td>0.06</td>
<td>0.81</td>
<td>0.00</td>
<td>0.13</td>
</tr>
<tr>
<td>Happiness</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Neutral</td>
<td>0.00</td>
<td>0.04</td>
<td>0.15</td>
<td>0.81</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Anger</th>
<th>Sadness</th>
<th>Happiness</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>0.95</td>
<td>0.00</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>Sadness</td>
<td>0.00</td>
<td>0.79</td>
<td>0.03</td>
<td>0.18</td>
</tr>
<tr>
<td>Happiness</td>
<td>0.02</td>
<td>0.00</td>
<td>0.91</td>
<td>0.08</td>
</tr>
<tr>
<td>Neutral</td>
<td>0.01</td>
<td>0.05</td>
<td>0.02</td>
<td>0.92</td>
</tr>
</tbody>
</table>

Busso et al, Analysis of emotion recognition using facial expressions, speech and multimodal information, ICMI, 2004
Is the child “certain”?

Trained decision trees using “leave-one-person-out” technique
- Used different subsets of features: lexical, acoustic, and visual
  - Lexical: “I don’t know”, false starts, repetitions, etc.
  - Acoustic: pauses, elongations, voice loudness, etc.
  - Visual: head/face/hand movements, facial expressions

<table>
<thead>
<tr>
<th>System</th>
<th>% Agreement *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline (always guess certain)</td>
<td>72.32</td>
</tr>
<tr>
<td>Lexical</td>
<td>73.41</td>
</tr>
<tr>
<td>Visual</td>
<td>74.17</td>
</tr>
<tr>
<td>Acoustic</td>
<td>82.66</td>
</tr>
<tr>
<td>Lexical + Visual</td>
<td>74.17</td>
</tr>
<tr>
<td>Acoustic + Visual</td>
<td>81.94</td>
</tr>
<tr>
<td>Acoustic + Lexical</td>
<td>82.29</td>
</tr>
<tr>
<td>Acoustic + Lexical + Visual</td>
<td>82.66</td>
</tr>
<tr>
<td>Average Human Agreement</td>
<td>86.15</td>
</tr>
</tbody>
</table>

All agreement statistics are pairwise agreement percentages with the ground truth.

Is the child “certain”?
Tracking Emotion Trends using Body Language and Speech

• Emotions continuously unfold with variable intensity and clarity
• Body language reflects the underlying emotion

Goal
  – Track continuous emotional trends through time
    • Emotion tracking instead of emotion recognition
  – Understand how body language is modulated by emotion

Approach/Contributions
  – Detailed body language descriptions
  – Joint modeling of audio-visual expressions and emotion
  – Tracking emotional trends at variable levels of detail

Results/Findings
  – Promising performance for activation and dominance trend tracking
  – Importance of body language features
    • Looking at/away, leaning forward/away, walk vs run etc

Body Language Feature Extraction

Front View

Back View

Person A

Person B

hand position (xh,yh,zh)

local system

x

y

z

global system

\( |V| \): absolute velocity

\( V' \): relative velocity of A towards B

Person A

Person B

angle

x

y

z
Framework Overview

Body Language Feature Extraction from MoCap data

Speech Feature Extraction from speech regions

Feature vector at time t

Features Xt Target Emotion Yt

GMM: P(X,Y)

GMM: P(Y|X)

Statistical Mapping from a new observed X to Y

\( \hat{Y} = \text{argmax } P(Y | X) \)

Collection of Audio, Video and MoCap data of dyadic improvisations

Annotation of Continuous Emotional Attributes from multiple people

Modeling and Estimation of Continuous Emotional Attributes
Some Applications
Behavior Coding: Distressed Couple interactions

10-minutes long problem solving interaction

Coding is performed at the session-level

Example coding goal:
“Is the husband showing acceptance?” (scale 1-9)

From the manual:
“Indicates understanding and acceptance of partner’s views, feelings, and behaviors. Listens to partner with an open mind and positive attitude. … ”

Husband speaking turns:

Jones and Christensen, Social Support Interaction Rating System, UCLA, 1998
Estimate behavioral codes from data

Multimodal Signals

Visual
- Head Orientation
- Body orientation
- Velocity of arms
- Openness of posture

Lexical
- Words
- Fragments
- Word boundaries
- Topic

Acoustic
- Voice Activity
- Pitch
- Energy
- Spectral
Focus on extreme cases of session-level judgments

Sample codes:
acceptance, blame, positive affect, negative affect, sadness, humor

“HOW MUCH DO TWO PEOPLE SYNCHRONIZE IN A CONVERSATION?”

Autism: Results on Atypical Prosody

Child’s Prosody

- “Monotone”
  p<0.01
- “Abnormal volume”
  p<0.05
- “Breathy/Rough”
  p<0.01
- Slower speaking rate
  p<0.05

Psychologist’s Prosody

- questions/affect
  p<0.05
- variable prosody
  p<0.01
- also higher jitter
  p<0.01
- slower/then faster
  p<0.01

The psychologists may be varying their engagement strategies

Child-Parent Physiological Synchrony

Overt signals of children with autism might be inconsistent with their inner affective state

**Electrodermal response (EDA): socio-cognitive load (arousal levels)**

**Verbal Response Latency (VRL): reflect cognitive and affective state**

**APPROACH**

Joint representation of child and parent physiological cues with coupled HMMs modeled on physiological features

Predict cognitive load with synchrony features

**FINDINGS**

Language and physiology give complementary information

Parent’s cues provide additional information about child’s behavior

Parents tend to synchronize with their children depending on the child’s ability to engage in task

---


Psychotherapy: Motivational Interviewing, for Addiction

Widely used in psychotherapy
Client's (interviewee) own will of making a change
Therapist (interviewer): understand, facilitate, do not dictate
Goal-oriented, highly-structured
Non-confrontational, non-judgmental, dialog setting

Computational behavior Modeling: POSSIBILITIES
Interaction dynamics of interviewer-interviewee: E.g., Empathy:
Computational language modeling provides useful insights into the expressed empathy behavior of therapists
Use speech, spoken language, nonverbals, body language cues
Data from several clinical intervention studies, coded by experts

Bo Xiao, Dogan Can, Panayiotis G. Georgiou, David Atkins and Shrikanth S. Narayanan. Analyzing the Language of Therapist Empathy in Motivational Interview based Psychotherapy. Proceedings of APSIPA 2012


TALK SUMMARY:

BSP Open Challenges → RICH R&D Opportunities

- Robust capture and processing of multimodal signals
- Capturing natural behavior in ecologically valid ways
- Behavior representations for computing
- Reflecting multiple (diverse) perspectives and subjectivity
- Feature-behavior correspondence: human like processing
- Scientifically and Computationally principled modeling
- Reliably characterizing atypical and disordered patterns
- Data provenance, integrity, sharing, and management

- Developing productive partnerships with domain experts
Behavioral signal processing

COMPUTING

OF human action and behavior data

FOR meaningful analysis: timely decision making & intervention (action)

BY collaborative integration of human expertise with automated machine processing

HUMANS
Concluding Remarks:
Enabling Behavioral informatics

• Human behavior can described from a variety of perspectives
  • Both challenges and opportunities for R&D
  • Multimodal data integral to derive and model these features
• Computational advances: sensing, processing and modeling
  • Support BOTH human and machine decision making
• Exciting technological and societal possibilities
  • Opportunities for interdisciplinary and collaborative scholarship

BEHAVIORAL signal processing:

✓ Help do things we know to do well more efficiently, consistently
✓ Help handle new data, create new models to offer unimagined insights
✓ create TOOLS for discovery

SUPPORTED BY: ONR, ARMY, DARPA, IARPA, NSF, NIH