Emotions in Engineering: Methods for the Interpretation of Ambiguous Emotional Content

Emily Mower
April 29, 2011
Motivation

• Increasing prevalence of interactive technology
  • Importance of emotion understanding

• Engineering research starting to overlap with human behavioral research:
  • Autism
  • Depression
  • Marital therapy
  • General interaction dynamics
  • Psychiatric disorders
**Motivating Example**

**Emotional Computer Assistant**

- Provides interaction assistance
- Describes the emotions of others
- Allows user to understand stimuli for proper response

**User:**

**Other:**

**Assistant:**

- "The other person is frustrated... this is a mix of anger and sadness"
- "I am sorry"
Focus of this Presentation

**Emotion Profiles**: A novel mid-level representation for quantifying emotion

- **Overview**:
  - Alleviates limitations of current frameworks
  - Captures shades of emotion
  - Represents ambiguous utterances
  - Component of classification
  - Stand-alone representation
  - Interpretable and informative
  - Can be used in a user-personalization framework

**Key finding**: EPs can be used to track the emotional trajectory of audio-visual utterances
Data Overview: USC IEMOCAP

• Data:
  • 5 m-f pairs of actors
  • *Audio*, video, motion-capture \((x,y,z)\)

• Elicitation Strategy:
  • Scripted sessions
  • Improvisation scenarios

• Emotional descriptors:
  • Categorical
  • Dimensional

*Data collection led by Carlos Busso, UT Dallas*
Feature Extraction and Selection

• Extraction:
  • Utterance-length
  • Mean, variance, range, upper-quantile, lower-quantile, quantile range

• Final feature set:
  • Principal Feature Analysis
  • Top 30 features

Audio Features:
Prosodic: pitch and energy
Spectral: Mel Filterbank Coefficients

Video Features:
Motion capture relative distances
Mouth, Eyebrows, Cheek, Forehead
Describe the presence or absence of multiple emotion classes in a single clip using an estimate of **classifier confidence**

- **Binary Support Vector Machine classifications**
  - Self vs. other
  - Matlab implementation

- **Output:**
  - Binary yes/no for class membership
  - Distance from hyperplane

- **Interpretation:**
  - Weight the binary output by the distance from the hyperplane (“confidence”)

Classification:
- Angry vs. Not Angry
- Happy vs. Not Happy
- Sad vs. Not Sad
- Neutral vs. Not Neutral
Emotion Profile Construction

Form semantic clusters using disjoint set of speakers

Train Self vs. Other Binary Classifiers on Each Semantic Cluster

4-Dimensional Profiles For Test Speaker

\[ d = \frac{t_m y(x_m)}{\|w\|} \]

\[ \sum_{n=1}^{N} a_n t_n e^{-\|x_m - x_n\|^2} + b \]

Use trained binary classifiers to create an estimate of the emotion content
Distance-Based Profile Measures

Valence-Activation plot for TOTAL AGREEMENT

- Angry
- Happy
Emotograms: Dynamic Emotion Profiles

Emotion Profile for an Utterance Labeled “Happy”

![Emotion Profile Graph]

Emotogram for an Utterance Labeled “Happy”

![Emotogram Image]
Goal: Classify the affective state of clips at the utterance level using Emotograms

- Features extracted over 10 (5m/5f) IEMOCAP speakers:
  - Motion capture: relative distances
  - Audio: prosodic, spectral
  - Feature Selection: Principal Feature Analysis (30 features)

- Extract EPs over window lengths: 0.25 – 2 seconds
  - Train binary angry, happy, neutral, sad SVMs on disjoint set of speakers (9)

- Model the trajectory of the EPs
  - Train angry, happy, neutral, sad HMMs on disjoint set of speakers (9)

- Validation:
  - Leave-one-subject-out cross-validation (over each test speaker, merged results)
Emotogram Construction
Results

Classification accuracy as a function of sentence length

Accuracy (%) vs Window length (s)

- 6 -- inf
- 3 -- 6
- 1.5 -- 3
- 0.5 -- 1.5
Conclusions and Future Directions

• Hierarchical system improves classification performance over all sentence lengths when compared to static only (absolute / relative):
  • 6+ – 7.84% / 11.75%
  • 3-6 – 3.55% / 5.48%
  • 1.5-3 – 0.54% / 0.87%

• Largest improvement with longest sentences:
  • Implies that there exists a recognized pattern of emotion fluctuation

• Human ability:
  • We can tell when emotions sound “wrong”
  • Flat affect is a diagnostic tool

• Implication:
  • Emotion modulations can be modeled by people
  • This modulation may be modeled using a grammar
Published Work in Emotion Profiles


Thanks!

Questions?