Confusion Mental State Inference through Intel RealSense

CS2951K Final Project Proposal

Ning Hou, Lee Painton, Eric Rosen

May 14, 2015

1 Research question

Affect display is the combination of facial, gestural and vocal cues by which persons consciously or unconsciously communicate emotion. Cues such as facial expression, vocal prosody and gestural display are all modes by which the affective state of an individual can be inferred. We are specifically interested in exploring the recognition of confusion in a person by means of their facial landmarks. By confusion we mean the common definition of a mental state where, given a situation, a person either understands it or is confused. We believe that there is a correlation between changes in a person's facial cues and the experience of confusion.

2 Significance

Reliably determining user affect is an open problem in HCI and part of a field called affective computing. The development of affect sensitive intelligent agents would allow computers to interact more effectively with humans in tasks where emotion has an impact, for example learning or driving. Confusion is especially significant during these tasks as it can actively interfere, or even be dangerous in the case of tasks such as operating heavy machinery. Detecting confusion can also be a valuable aid in the diagnoses of medical conditions which may not be immediately obvious to the human observer. Confusion also serves intuitively as a natural perceptual feedback respresenting the efficacy of an intelligent agent's communication attempts and could be incorporated as part of the reward function in a learning agent. It is our more immediate hope that we can utilize work in this project to make a Baxter robot aware of confusion in subjects with whom it is interacting.

3 Methodology

In work establishing a framework for machine Emotional Intelligence, Picard et all [?] discuss factors which need to be considered when gathering data for experimental purposes. For our experiment we are interested primarily in event-elicited emotions which arise unconsciously based on the situation. To this end we have designed a quiz of five questions which are intended to provide a spectrum of data. During the course of each question we collect a set of datapoints using an Intel RealSense device which we have attached to a computer and pointed at the quizee. This set includes 13 facial landmarks and 10 emotional features. The facial landmarks are as follows:

- Left eyebrow raiser
- Right eyebrow raiser
- Left eyebrow lowerer
- Right eyebrow lowerer
- Mouth open
- Mouth smile
- Mouth kiss
- Left eye closed
- Right eye closed
- Eyes turn left
- Eyes turn right
- Eyes turn up
- Eyes turn down

3.1 Confusion quiz

To capture the human facial and pulse response to mental state of confusion, we design a short quiz on a scale of questions from easy (less confusing) to hard (more confusing):

1. *What is your name?*
   - Use: Calibrate neutral features.
2. *Who is the President of the United States?*
   - Answer: Barak Obama
   - Use: Easy question measures non-confusing features.
3. *How many fingers am I holding up? (Hold up four)*
   - Answer: Four
   - Use: Easy question measures non-confusing features.
4. *I have two coins totaling 15 cents, one of which is not a nickle. What are the two coins?*
   - Answer: A dime and a nickle
   - Use: Medium question that might seem confusing at first but can be answered after some thinking or clarification. This question measures both confusing and non-confusing features, as well as the transition.
5. *Has anyone really been far enough and decided to use even what they look like?*
   - Answer: Nonsense
   - Use: Intentionally confusing question to measure confusing features.
6. *Theres a dead man in a room surrounded by 53 bicycles. Why is he dead?*
   - Answer: He was caught cheating at cards.
   - Use: Intentionally confusing riddle to measure confusing features.
7. *Make a face of confusion.*
8. *Make a face of understanding.*
   - Use: Extra features of acted features of confusion and non-confusion (understanding).

3.2 Dataset

The dataset consists of two components to the confusion quiz questions in 3.1:

1. The video of the face and upper body of the test subject during the quiz process.
    - Three authors independently annotate the video frames by label of confusion and non-confusion.
    - The intersection of annotated confusion frames forms the **baseline** for confusion evaluation and inference.
2. The features detected by Intel RealSense built-in face tracking and emotion modules at 0.2 millisecond (ms) frame:
    - Pulse [in beats per minute (BPM)]
    - Facial landmarks [on a scale of 0 to 100]:
        - Brow: Raise left [AU1,2], Raise right [AU1,2], Lower left [AU4], Lower right [AU4]
        - Mouth: Smile, Kiss, Open
        - Head: Turn left [AU51], Turn right [AU52], Up [AU53], Down [AU54], Tilt left [AU/M55], Tilt right [AU/M56]
        - Eyes: Turn left [AU/M61], Turn right [AU/M62], Up [AU63], Down [AU64]
    - Emotions [on a scale of -1 to 10]:
        - Primary: Anger, Contempt, Disgust, Fear, Joy, Sadness, Surprise
        - Sentiments: Negative, Positive, Neutral

The facial landmarks that fit into the definition of action units have been denoted with [AU]. We could also consider the initial frames of neutral face as AU0. (Because mouth features do not overlap with AU definitions, we only use the feature names in our method. However, we the AU notations here for future use and comparison with other facial action and emotion research.)

Additionally, the datasets were collected in the same format under two scenarios:

1. Conversation: we asked the quiz questions to human subjects.
2. Computerized test: the human subjects took the quiz on computer.

3.3 Naive Bayes

Considert the features conditionally independent and frames taken at every 0.2-ms timestamp independent inputs. We formulate the Naive Bayes classifier for confusion mental states: given the feature vector $X = x_1, ..., x_{28}$ consisting of the 28 features described in Section 3.2.2, we compute $P(\text{confusion}|X)$ by Bayes Rule:

$$P(\text{confusion}|X) = \frac{P(\text{confusion})P(X|\text{confusion})}{P(X)} \tag{1}$$

$$\propto P(\text{confusion})P(X|\text{confusion}) \tag{2}$$

Assuming conditional independence for features in X,

$$P(\text{confusion}|X) \propto P(\text{confusion}) \prod_{i=1}^{28} P(x_i|\text{confusion}) \tag{3}$$

where

$$P(\text{confusion}) = \frac{\text{count of confusion frames}}{\text{total number of frames}} \tag{4}$$

$$P(x_i|\text{confusion}) = \frac{\text{count of feature } x_i \text{ in confusion class}}{\text{total number of features in confusion class}} \tag{5}$$

We plan to take questions $1, 5, 7, 8$ of Section 3.1 as the training data to compute $P(\text{confusion}|X)$ and evaluate the performance on questions $2, 3, 4, 6$ as testing data.

3.4 Bayes Filter

4 Results

4.1 Baseline

We form the baseline of confusion by taking the intersection of three independent set of annotated frames. The annotation is based on the

Table: Annotated frames labelled as confusion in our baseline

Subject	Confusion frames
DK	
John	
Nakul	
Paige	

4.2 Naive Bayes

Table: Discrimitive weights of confusion in each feature

		Baseline	Naive Bayes	Bayes Filter
Pulse				
Brow	Raise left			
Brow	Raise right			
Brow	Lower left			
Brow	Lower right			
Mouth	Smile			
Mouth	Kiss			
Mouth	Open			
Mouth	Smile			
...				

5 Schedule

Date	**TODO**
2/26 - 3/5	Finalize theoretic framework and experiment design
3/6 - 3/12	Program initial models and test with false data
3/13 - 3/19	Design interview script and post interview survey. Find interview subjects and schedule
3/20 - 3/26	Have at least 10 subjects interviewed with collected data or move to backup plan
3/27 - 4/2	Test data on models and compare
4/2 - 4/7	Checkpoint presentation
4/8 - 4/14	Collect more data as needed. Adjust and formalize the model based on results
4/15 - 4/21	Prepare results
4/23 - 4/28	Final presentation

6 Tables and Figures

Figure 1: Sample dataset taken at one 0.2 ms timestamp

References

[1] Ashish Kapoor, Selene Mota, and Rosalind W Picard. Towards a learning companion that recognizes affect. In *AAAI Fall symposium*, pages 2–4, 2001.