Third International Workshop on Context Based Affect Recognition
CBAR 2015

Person-specific behavioural features for automatic stress detection

Jonathan Aigrain - Séverine Dubuisson
Marcin Detyniecki - Mohamed Chetouani

May 4, 2015
Understanding nonverbal cues and social signals is one of the key elements in human-human interaction

- Nonverbal communication conveys an important part of the meaning
- Interpretation of these signals is person and context dependent

**Context**: stressful situations

**Objective**: Automatically detect stress in an unobtrusive way by studying the body language
Stress definition

Definition proposed by Lazarus [4] with the conditions regarding the stimulus proposed by Koolhaas et al. [3]

- Stress is the result of a transaction between a person and her environment
- This transaction includes:
  - A stimulus considered as uncontrollable and/or unpredictable
  - An evaluation of the situation and the conclusion of the presence of a threat
  - Coping processes
  - Several effects on mind and body
Related works

Automatic stress detection techniques are mainly using:
- Speech signals
- Physiological signals

Few works address stress detection using body language:
- Giakoumis et al. [2] enhanced the performance of stress detection systems with behavioural features
- Soury [7] used postural features in a multimodal fusion model
- Lefter [5] used visual features such as HOG and HOF to predict intermediate level variables, which are then used to predict stress
Presentation outline

- Data collection
- Feature extraction
- Person-specific normalization
- Stress detection
- Conclusions
According to Dickerson and Kemeny [1], there are 4 main classes of stressors:

- Cognitive tasks
- Public speaking
- Noise exposure
- Emotion induction

Experiments which combine public speaking and cognitive tasks are considered the most effective
Stress induction procedure

The experiment used is an evaluated time-constrained mental arithmetic test:

- 6 steps of increasing difficulty
- Performed in front of two people
- Biased performance feedback
Acquired data

For each of the 14 participants, for each of the 6 steps:

- Video of the whole body in 640 x 480 from the Kinect
- Skeleton from the Kinect
- Video of the face in 1440 x 1080 from the HD camera
- Self-assessed stress level using a Likert-scale (1-5)
Feature extraction

Body language features

- Quantity of Movement
  - Computed from the skeleton and from the image
  - Skeleton QoM computed for the head
  - FFT applied on the image QoM, divided in 10 bins
- Detection of periods of high activity
  - Using the peaks of the QoM
Feature extraction

Body language features

- Detection of posture changes

- Detection of self-touching
  - Self-touching in the region of the head
  - Fingers rubbing

25 features
Facial features

- Activation level of 12 Action Units
  - AU are presented by Ekman in the Facial Action Coding System
  - Extracted using the method of Nicolle et al. [6]
  - Average and standard deviation used as features

24 features
Objective: Reducing the impact of interindividual differences
Hypothesis: Stress is easier to detect if we look at the evolution of one’s behaviour

\[ \tilde{f}_{ps} = \frac{f_{ps} - f_{p1}}{f_{p1}} \]

with \( f_{ps} \) the vector of features for the person \( p \) on step \( s \) and \( \tilde{f}_{pj} \) the normalized vector
Evaluation process

• Classification using SVM with three kernel functions
  • Linear
  • Radial Basis Function
  • Polynomial
• Each video is associated with a label
  • Stress (S) if self-assessed stress level $> 3$
  • Non-Stress (NS) otherwise
• Leave-One-Subject-Out cross-validation
• Mean accuracy over 10 runs
Results

<table>
<thead>
<tr>
<th>kernel type</th>
<th>raw</th>
<th>normalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poly</td>
<td>0.64 ± 0.04</td>
<td>0.77 ± 0.02</td>
</tr>
<tr>
<td>RBF</td>
<td>0.65 ± 0.03</td>
<td>0.76 ± 0.02</td>
</tr>
<tr>
<td>Linear</td>
<td>0.67 ± 0.01</td>
<td>0.77 ± 0.01</td>
</tr>
</tbody>
</table>

- No significant difference between the 3 kernel functions
- Person-specific normalization improves accuracy:
  - Poly: +20%
  - RBF: +17%
  - Linear: +15%
Impact of the features set

<table>
<thead>
<tr>
<th>features set</th>
<th>raw</th>
<th>normalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0.67 ± 0.01</td>
<td>0.77 ± 0.01</td>
</tr>
<tr>
<td>Face</td>
<td>0.68 ± 0.01</td>
<td>0.65 ± 0.03</td>
</tr>
<tr>
<td>Body</td>
<td>0.63 ± 0.03</td>
<td>0.80 ± 0.01</td>
</tr>
</tbody>
</table>

- Results obtained with the linear kernel
- Person-specific normalization effective only on body features (+27%)
- Raw facial features give better results than raw body features
  - May be explained by less interindividual differences in facial expression
Performances of individual features

- Classification obtained with only one feature
- SVM with RBF kernel function
  - Allows several “split values” along the feature axis
- Leave-One-Subject-Out cross-validation
- Mean accuracy over 10 runs
- 5 best and 5 worst features are presented
Performances of individual raw features

<table>
<thead>
<tr>
<th>feature</th>
<th>accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCC</td>
<td>0.73 ± 0.02</td>
</tr>
<tr>
<td>AU9 - std</td>
<td>0.72 ± 0.01</td>
</tr>
<tr>
<td>FFT2</td>
<td>0.72 ± 0.01</td>
</tr>
<tr>
<td>AU4 - std</td>
<td>0.72 ± 0.01</td>
</tr>
<tr>
<td>AU4 - mean</td>
<td>0.72 ± 0.01</td>
</tr>
<tr>
<td>AU25 - mean</td>
<td>0.60 ± 0.03</td>
</tr>
<tr>
<td>RHM</td>
<td>0.58 ± 0.01</td>
</tr>
<tr>
<td>AU5 - std</td>
<td>0.58 ± 0.02</td>
</tr>
<tr>
<td>AU9 - mean</td>
<td>0.56 ± 0.01</td>
</tr>
<tr>
<td>HAPMD</td>
<td>0.55 ± 0.04</td>
</tr>
</tbody>
</table>

- Better performances than the whole set of features (67%)
- Good results for posture changes (PCC) and brows activity (AU4 and AU9)
- Difficult to interpret what FFT2 means
Performances of individual normalized features

<table>
<thead>
<tr>
<th>feature</th>
<th>accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFT1</td>
<td>0.76 ± 0.02</td>
</tr>
<tr>
<td>HAPC</td>
<td>0.74 ± 0.01</td>
</tr>
<tr>
<td>FFT7</td>
<td>0.73 ± 0.02</td>
</tr>
<tr>
<td>HAPMV</td>
<td>0.73 ± 0.02</td>
</tr>
<tr>
<td>PCC</td>
<td>0.73 ± 0.02</td>
</tr>
<tr>
<td>AU1 - std</td>
<td>0.49 ± 0.03</td>
</tr>
<tr>
<td>AU25 - mean</td>
<td>0.46 ± 0.02</td>
</tr>
<tr>
<td>AU26 - mean</td>
<td>0.45 ± 0.02</td>
</tr>
<tr>
<td>AU15 - mean</td>
<td>0.45 ± 0.03</td>
</tr>
<tr>
<td>AU17 - std</td>
<td>0.43 ± 0.04</td>
</tr>
</tbody>
</table>

- Similar performances than the whole set of features (77%)
- Normalization effective only on body features
- Good results for body activity (FFT1), periods of high activity (HAPC and HAPMV) and posture changes (PCC)
- Difficult to interpret what FFT7 means
Conclusions and possible improvements

Conclusions

• Unobtrusive solution for stress detection
• Person-specific normalization effective only on body language features
• Using only one feature can provide good classification accuracy
  • 73% for number of posture changes
  • 76% for normalized 1st bin of the FFT

Improvements

• Feature selection
• Using normalized body features and raw facial features
Thank you for your attention!
Bibliography


