Personalized Automatic Estimation of Self-reported Pain Intensity from Facial Expressions

Daniel Lopez-Martinez¹,² (dlmocdm@mit.edu)
Ognjen (Oggi) Rudovic¹
Rosalind Picard¹

¹ Affective Computing group @ MIT Media Lab
² Harvard-MIT Division of Health Sciences and Technology
What is pain?

“Pain is a distressing experience associated with actual or potential tissue damage with sensory, emotional, cognitive and social components.”

Affective: Negative emotion: anxiety, fear, unpleasant sensation.

Sensory: Perception of pain characteristics: intensity, quality, location.

Cognitive: Interpretation of pain.

Pain ≠ nociception

- Nociception refers to the peripheral and central nervous systems processing information generated by stimulation of nociceptors by noxious stimuli.
- Nociception can occur in the absence of pain.

- Pain is a product of higher brain center processing of signals it has received.
- Pain can occur in the absence of nociception (e.g. neurogenic pain).

Figures reproduced from: Wall & Melzack's Textbook of Pain, Sixth Edition
A conceptual framework for understanding pain

Self-report: the gold standard of pain measurement

Pain descriptors based on intensity ratings by patients

Wall & Melzack's Textbook of Pain, Sixth Edition
Self-report: the gold standard of pain measurement

**VAS**
Visual Analog Scale

0
No pain

10
Worst possible pain

**NRS**
Numerical Rating Scale

0 1 2 3 4 5 6 7 8 9 10
No pain Moderate pain Unbearable pain
When self-report fails... Automatic Pain Recognition

When pain cannot be communicated

Large-scale clinical studies

Human-robot interactions
Pain Recognition using Facial Expressions: The Prkachin and Solomon Pain Intensity (PSPI) metric

- **Facial Action Coding System (FACS)** is a system to taxonomize human facial movements by their appearance on the face.
- Each observable component of facial movement is called an *Action Unit* or AU. There are 98 AUs.
- AUs are manually coded, for each frame.
- The *Prkachin and Solomon Pain Intensity (PSPI)* score is a metric that measures pain as a linear combination of the intensities of facial action units. It is defined on an ordinal scale 0-15.

\[
\text{PSPI} = \text{AU4} + \max(\text{AU6},\text{AU7}) + \max(\text{AU9},\text{AU10}) + \text{AU43}
\]

**Example of action units (AUs)**

- Inner Brow Raiser
- Outer Brow Raiser
- Brow Lowerer
- Upper Lid Raiser
- Cheek Raiser
- Lid Tightener
- Lid Droop
- Slit
- Eyes Closed
- Squint
- Blink
- Wink
- Nose Wrinkler
- Upper Lip Raiser
- Nasolabial Deepener
- Lip Corner Puller
- Cheek Puffer
- Dimpler
- AU 15
- AU 16
- AU 17
- AU 18
- AU 20
- AU 22
- Lip Corner Depressor
- Lower Lip Depressor
- Chin Raiser
- Lip Puckerer
- Lip Stretchcer
- Lip Funneler
- AU 23
- AU 24
- *AU 25
- *AU 26
- *AU 27
- AU 28
- Lip Tightener
- Lip Pressor
- Lips Part
- Jaw Drop
- Mouth Stretch
- Lip Suck

- Brow lowerer
- Cheek raiser
- Lid tightener
- Upper lip raiser
- Eyes closed
From facial expressions to VAS scores

Facial AUs and their intensity levels used to derive PSPI (=12) for target face image.

Facial landmarks obtained using an Active Appearance Model.

Facial Expressions

Individual Facial Expressiveness Score (I-FES)

How do we go from PSPI scores to VAS?
Individual Facial Expressiveness Score (I-FES)

The Individual Facial Expressiveness Score (I-FES) captures the ratio of OPI, obtained by independent observers, to VAS, self-reported by the subject.

\[
p_i = \begin{cases} 
\frac{1}{\alpha} \sum_{k=1}^{\alpha} \frac{o_i^k+1}{v_i^k+1}, & \text{iff } \alpha > 0 \\
1, & \text{otherwise}
\end{cases}
\]

- \(i\): subject number
- \(p_i\): I-FES for subject \(i\)
- \(k\): sequence number for subject \(i\)
- \(\alpha\): number of sequences for subject \(i\)
- \(o_i^k\): observed pain intensity (OPI) for subject \(i\), sequence \(k\)
- \(v_i^k\): VAS rating for subject \(i\), sequence \(k\)

The VAS vs OPI scores for four different subjects (along with the number of available sequences).
pRNN-HCRF model for personalized VAS estimation

Recurrent neural network
*Input*: facial landmarks for each frame in the window
*Output*: predicted PSPI score for the central frame in the window

Facial landmark coordinates obtained using an Active Appearance Model

Hidden Conditional Random Field
*Input*: predicted PSPI scores for all frames in the sequence + I-FES score
*Output*: predicted VAS score for the whole sequence

Individual Facial Expressiveness Score (I-FES)
Long Short-Term Memory (LSTM) Neural Networks

**Input**
- Location of 66 facial landmarks/frame obtained with an active appearance model, for 8-8 frames (17 frames in total).
- Applied PCA, resulting in 40D feature vectors per frame (preserving 95% of variance).
- Standardized input.

**Output**
- Predicted PSPI Score for each 8-8 window.

LSTM figure reproduced from: http://colah.github.io/posts/2015-08-Understanding-LSTMs
An HCRF models the conditional probability of a class label given a set of observations by:

$$ P(y|x, \theta) = \frac{\sum_s P(y, s | x, \theta)}{\sum_{y' \in Y, s \in S^m} e^{\Psi(y', s, x; \theta)}} $$

Objective function:

$$ L(\theta) = \sum_i \log P(y_i | x_i, \theta) - \frac{1}{2\sigma^2} ||\theta||^2 $$

**Input:** Predicted PSPI scores for each window in the sequence  
**Output:** Predicted VAS score
pRNN-HCRF model for personalized VAS estimation

Algorithm 1 Personalized RNN-HCRFs

Learning: Input $\mathcal{D}_{tr} = \{X_i^{tr}, Y_i^{tr}, O_i^{tr}, S_i^{tr}\}_{i=1}^{L_{tr}}$

Step 1: Optimize RNNs ($Q^{opt}$) given $\{S_{tr}, X_{tr}\}$

Step 2: Compute I-FES

for $i = 1: L_{tr}$, $p_i \leftarrow \{Y_i^{tr}, O_i^{tr}\}$ end, $\mathcal{P} = \{p_i\}_{i=1}^{L_{tr}}$

Step 3: Optimize HCRFs ($\Omega$)

a) estimate PSPI $\tilde{S} \leftarrow \text{RNN}(X_{tr}; Q^{opt})$ \text{--- Step 1}, $S_p = \{\tilde{S}, \mathcal{P}\}$

b) $\min_{\Omega^{opt}} \sum_{i=1}^{L_{tr}} \sum_{j=1}^{N_i} \log P(v_i^j | S_i^j; \Omega) + \lambda ||\Omega||^2$

Output: RNNs($\cdot$; $Q^{opt}$), HCRFs($\cdot$; $\Omega^{opt}$)

Inference: Input $\mathcal{D}_{te} = \{X_{*}, p_{*}\}$

Step 1: Estimate PSPI $S^* \leftarrow \text{RNNs}(X_{*}; Q^{opt})$

Step 2: Estimate VAS $v^* \leftarrow \text{HCRFs}(S_p^* = \{S^*, p^*\}; \Omega^{opt})$

Output: $S^*, v^*$
UNBC-McMaster Shoulder Pain Expression Archive Database

25 subjects
200 sequences
48,398 frames (16.4% with pain)

PSPI prediction performance

The mean MAE/ICC(3,1) and standard deviation (for 5 repetitions) for the LSTM-RNN, a neural network (NN) with one hidden layer and 200 hidden units, and SVR (C = 0.1, ε = 0.01), for the predictions of PSPI from facial landmarks, computed on test persons.

Confusion matrix for the predicted PSPI scores using the RNN-LSTM.
VAS prediction performance (1/2)

The performance of different methods tested for VAS (0-10) estimation. The mean and standard deviation are computed over 5 random selection of target sequences used to compute I-FES for test persons.
Confusion matrices for the predicted VAS scores using the non-personalized ($\alpha=0$) and personalized ($\alpha=1,2$) RNN-HCRF model.
Any questions?

Daniel Lopez-Martinez

www.daniellopez.eu
dlmocdm@mit.edu