Machines' Learning on Social Behaviour Sensing

Björn W. Schuller
Machine Learning Group
Imperial College London
Chair of Complex & Intelligent Systems
University of Passau
audEERING UG
<table>
<thead>
<tr>
<th>Year</th>
<th>Class</th>
<th># Classes</th>
<th>%UA/*AUC/*CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>Nativeness</td>
<td>[0,1]</td>
<td>43.3*</td>
</tr>
<tr>
<td></td>
<td>Parkinson’s</td>
<td>[0,100]</td>
<td>54.0*</td>
</tr>
<tr>
<td></td>
<td>Eating</td>
<td>7</td>
<td>62.7</td>
</tr>
<tr>
<td>2014</td>
<td>Cognitive Load</td>
<td>3</td>
<td>61.6</td>
</tr>
<tr>
<td></td>
<td>Physical Load</td>
<td>2</td>
<td>71.9</td>
</tr>
<tr>
<td>2013</td>
<td>Social Signals</td>
<td>2x2</td>
<td>92.7*</td>
</tr>
<tr>
<td></td>
<td>Conflict</td>
<td>2</td>
<td>85.9</td>
</tr>
<tr>
<td></td>
<td>Emotion</td>
<td>12</td>
<td>46.1</td>
</tr>
<tr>
<td></td>
<td>Autism</td>
<td>4</td>
<td>69.4</td>
</tr>
<tr>
<td>2012</td>
<td>Personality</td>
<td>5x2</td>
<td>70.4</td>
</tr>
<tr>
<td></td>
<td>Likability</td>
<td>2</td>
<td>68.7</td>
</tr>
<tr>
<td></td>
<td>Intelligibility</td>
<td>2</td>
<td>76.8</td>
</tr>
<tr>
<td>2011</td>
<td>Intoxication</td>
<td>2</td>
<td>72.2</td>
</tr>
<tr>
<td></td>
<td>Sleepiness</td>
<td>2</td>
<td>72.5</td>
</tr>
<tr>
<td>2010</td>
<td>Age</td>
<td>4</td>
<td>53.6</td>
</tr>
<tr>
<td></td>
<td>Gender</td>
<td>3</td>
<td>85.7</td>
</tr>
<tr>
<td></td>
<td>Interest</td>
<td>[-1,1]</td>
<td>42.8*</td>
</tr>
<tr>
<td>2009</td>
<td>Emotion</td>
<td>5</td>
<td>44.0</td>
</tr>
<tr>
<td></td>
<td>Negativity</td>
<td>2</td>
<td>71.2</td>
</tr>
</tbody>
</table>

**Affect.**

- **AVEC 2011 & 2012: Emotion**
  AVEC database: 50k Turns

- **AVEC 2013 & 2014: Depression**
  AVDLC database: 240 h, ~300 subjs.
  Beck Depression Inventory: 0 – 63

- **AV+EC 2015: Emotion**
  RECOLA database: 6 h, ~30 subjs.


Personality.

- MAPTRAITS: Perceived Perso.
  SEMAINE:
  44 clips, 11 subjects, 6 raters

<table>
<thead>
<tr>
<th>Agreement</th>
<th>Audio</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>Openness</td>
<td>.53</td>
<td>.39</td>
</tr>
<tr>
<td>Conscientious</td>
<td>.51</td>
<td>.28</td>
</tr>
<tr>
<td>Extroversion</td>
<td>.66</td>
<td>.38</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>.50</td>
<td>.41</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>.77</td>
<td>.35</td>
</tr>
<tr>
<td>Engagement</td>
<td>.65</td>
<td>.45</td>
</tr>
<tr>
<td>Likeability</td>
<td>.56</td>
<td>.37</td>
</tr>
<tr>
<td>Facial Attractivity</td>
<td>-</td>
<td>.34</td>
</tr>
<tr>
<td>Vocal Attractivity</td>
<td>.41</td>
<td>-</td>
</tr>
</tbody>
</table>

Working, Yet?
"YouTube Movie Reviews: In, Cross, and Open-domain Sentiment Analysis in an Audiovisual Context"

Eye-Contact.

- **Eye Contact from Acoustics**
  GRAS² Corpus: 28 + 4 subjects

"The acoustics of eye contact - Detecting visual attention from conversational audio cues", ACM Gaze-In, 2013.

<table>
<thead>
<tr>
<th>UA [%] / AUC</th>
<th>67.4 / .732</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eye-Contact</td>
<td></td>
</tr>
</tbody>
</table>
Social Traits.

- **Perceived Leadership Traits**
  409 recordings, 10 raters
  Acoustic & Linguistic analysis

<table>
<thead>
<tr>
<th>Trait</th>
<th>UA [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Malicious</td>
<td>61.3</td>
</tr>
<tr>
<td>Diplomatic</td>
<td>58.5</td>
</tr>
<tr>
<td>Achiever</td>
<td>72.5</td>
</tr>
<tr>
<td>Charismatic</td>
<td>63.2</td>
</tr>
<tr>
<td>Teamplayer</td>
<td>65.2</td>
</tr>
</tbody>
</table>

"The Voice of Leadership: Models and Performances of Automatic Analysis in On-Line Speeches",
Eating.

- **Speech Under Eating & Food**
  - 30 subjects, 6 food types, +ASR features

```
No  x
Ha  x  x
Ba  x  x
Ne  x  x
Ap  x  x
Cr  x  x
Bi  x  x
```

```
<table>
<thead>
<tr>
<th>R²</th>
<th>Crispness</th>
<th>.562</th>
</tr>
</thead>
</table>
```

**Action Units.**

- **Face Reading from Speech**
  GEMEP corpus, SVM vs DLSTMNNs, big (65) vs small (28) feature set

<table>
<thead>
<tr>
<th>AU</th>
<th>Onset</th>
<th>Apex</th>
<th>Offset</th>
<th>Occu.</th>
<th>Onset</th>
<th>Apex</th>
<th>Offset</th>
<th>Occu.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>61.88</td>
<td>64.54</td>
<td>59.48</td>
<td>67.03</td>
<td>62.08</td>
<td>67.96</td>
<td>61.18</td>
<td>67.62</td>
</tr>
<tr>
<td>2</td>
<td>64.38</td>
<td>67.43</td>
<td>62.27</td>
<td>69.33</td>
<td>64.60</td>
<td>70.99</td>
<td>62.98</td>
<td>70.90</td>
</tr>
<tr>
<td>4</td>
<td>61.35</td>
<td>66.21</td>
<td>63.33</td>
<td>64.66</td>
<td>61.90</td>
<td>66.67</td>
<td>63.19</td>
<td>67.05</td>
</tr>
<tr>
<td>6</td>
<td>64.01</td>
<td>65.17</td>
<td>63.35</td>
<td>63.78</td>
<td>64.25</td>
<td>67.71</td>
<td>63.17</td>
<td>63.11</td>
</tr>
<tr>
<td>7</td>
<td>62.71</td>
<td>56.62</td>
<td>58.29</td>
<td>54.47</td>
<td>62.85</td>
<td>59.74</td>
<td>61.46</td>
<td>52.72</td>
</tr>
<tr>
<td>10</td>
<td>61.82</td>
<td>60.53</td>
<td>56.39</td>
<td>60.39</td>
<td>61.92</td>
<td>61.03</td>
<td>58.02</td>
<td>60.34</td>
</tr>
<tr>
<td>12</td>
<td>59.34</td>
<td>59.85</td>
<td>56.71</td>
<td>58.67</td>
<td>58.04</td>
<td>60.92</td>
<td>57.77</td>
<td>58.72</td>
</tr>
<tr>
<td>17</td>
<td>60.47</td>
<td>64.78</td>
<td>62.34</td>
<td>64.53</td>
<td>61.62</td>
<td>64.94</td>
<td>63.13</td>
<td>65.88</td>
</tr>
<tr>
<td>Avg.</td>
<td>62.00</td>
<td>63.14</td>
<td>60.27</td>
<td>62.86</td>
<td>62.16</td>
<td><strong>65.00</strong></td>
<td>61.36</td>
<td>63.29</td>
</tr>
</tbody>
</table>
There is just one Vocal Production Mechanism…

- **Multiple-Targets**
  - Drunk
  - Angry
  - Has a Cold
  - Neurotic
  - Tired
  - Has Parkinson’s
  - Is Older

<table>
<thead>
<tr>
<th>% UA</th>
<th>Single</th>
<th>Multiple</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likability</td>
<td>59.1</td>
<td>(+A,G,Cl) 62.2</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>62.9</td>
<td>(+G,OCEA, Cl) 67.5</td>
</tr>
</tbody>
</table>
Signal Capture

Principle.

“Big Data vs. Little Labels”

Efficient Weakly Supervised

\[ s[k] \]

Learning at Signal’s Edge

Preprocessing

Feature Extraction

Learning

\[ y \]

/Decision

 UILabel

(e.g., “burst”)
Feature Extraction.

- **openSMILE**
  Brute-force High-Dim. Spaces
  (Android / C++)
  Online update

<table>
<thead>
<tr>
<th>#features</th>
<th>RTF</th>
</tr>
</thead>
<tbody>
<tr>
<td>10k</td>
<td>2%</td>
</tr>
<tr>
<td>500k</td>
<td>3%</td>
</tr>
</tbody>
</table>

```
Sensor-Signal
- Energy
- Harmonicity
- Fund. Freq.
- ...
- ...
- TF-Transform

Deriving | Filtering | Chunking | Deriving | Filtering |
---------|-----------|----------|----------|-----------|
Extremes  | Moments   | Peaks    | Segments | Spectral  |
Moments   | Peaks     | Segments | Spectral | Regression|
Peaks     | Segments  | Spectral | Regression|          |
Segments  | Spectral  | Regression|          |          |
Spectral  | Regression|          |          |          |
Regression|          |          |          |          |
```

“Recent Developments in openSMILE, the Open-Source Multimedia Feature Extractor”, ACM Multimedia, 2013.
(2nd place ACM MM Open Source Software Competition in 2010 and 2013, ~600 citations for 3 papers)
Less is More?

- **Geneva Minimalistic Set**
  
  **GeMAPS**: 18 LLDs / 62 functionals  
  Pitch, Jitter, Formant 1-3, Formant 1 bandwidth  
  Shimmer, Loudness, HNR  
  Alpha Ratio, Hammarberg Index, Spectral Slope 0–500Hz  
  and 500–1500 Hz, Formant 1-3 relative energy,  
  Harmonic difference H1–H2, Harmonic difference H1–A3

- **Extended**: +7 LLDs / 88 functionals  
  MFCC 1–4, Spectral Flux, Formant 2–3 bandwidth

Bag-of-X-Words

• **Audio Words**
  Base: UA/WA = 54.3/61.2%

VQ vs SVQ

**VAM**: Valence

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td># SV</td>
<td>SVCS [bit]</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
</tr>
</tbody>
</table>

Deep Learning.

- **Deep Neural Networks**
  
  DNN = MLP…

  Training makes the difference:

  Build them layer-wise
  
  Less uninitialized parameters

  Use “raw” features as input
  
  Net learns own higher-level features

  Use more training data

  On-line learning (stoch. gradient descent)

  Use ReLUs, drop-out learning, …

\[
y = f_{DNN}(x) = o \left( w^K \sigma \left( w^{K-1} \sigma (\ldots w^1 x) \right) \right)
\]

\[
E_{Tr}(y, y^*) = \sum_{x \in Tr} D(f_{DNN}(x), y^*)
\]

\[
w^{(i+1)} = w^{(i)} - \eta \frac{\partial E_B}{\partial w} (w^{(i)}), B \subset Tr
\]
Deep LSTM Nets.

• LSTM Cell
  
  Linear Unit
  Auto weight 1
  “Error Carousel”

  Non-linear Gate
  Input / Output / Forget

  Multiplicative
  Open / Shutdown

Example: ComParE: SSC

<table>
<thead>
<tr>
<th>% AUC (Acc.)</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN</td>
<td>83.3</td>
</tr>
<tr>
<td>Deep NN</td>
<td>92.4</td>
</tr>
<tr>
<td>LSTM</td>
<td>93.0</td>
</tr>
<tr>
<td>Deep LSTM</td>
<td>94.0</td>
</tr>
</tbody>
</table>

“Social Signal Classification Using Deep BLSTM Recurrent Neural Networks”, IEEE ICASSP, 2014. (Best Result ComParE: SSC)
Deep Recurrent Nets.

- “Very Deep” NN

\[ y_t = f_{RNN}(x_0, \ldots, x_t) = \sigma \left( W^o \sigma \left( Wx_t + R \sigma (\ldots Wx_0 + Rh_0) \right) \right) \]
Parallel Learning.

- **CURRENT**

  10 – 1k LSTM cells, 2k – 4Mio parameters, Distribution

Weakly Supervised Learning.

- **Transfer Learning**
  Re-use data of related domain

- **Active Learning**
  Select “most informative” instances from large amount of unlabelled data

- **Semi-Supervised Learning**
  Have computer label the data

- **Cooperative Learning**
  Efficient combination of the above
Transfer Learning.

- **Transfer Learning**
  Sparse Auto-Encoder
  target values \( \hat{x} \) input

<table>
<thead>
<tr>
<th>% UA / CC</th>
<th>Target</th>
<th>w/o</th>
<th>DAE</th>
<th>DAE-NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ComParE : EC</td>
<td>60.4</td>
<td>56.3</td>
<td>59.2</td>
<td>64.2</td>
</tr>
<tr>
<td>M ( \rightarrow ) S: A</td>
<td>.82</td>
<td>.05</td>
<td>.32</td>
<td>-</td>
</tr>
<tr>
<td>M ( \rightarrow ) S: V</td>
<td>.51</td>
<td>.06</td>
<td>.16</td>
<td>-</td>
</tr>
</tbody>
</table>


Cooperative Learning.

Stop Criteria:
- No data ‘likely’ sparse class
- Accuracy saturation

Confidence & Sparseness
- Sparse: AL
- Else: Discard
- High: SSL
- Else: AL

Model
- Final model

Classify
- Unlabelled instances

Train
- Labelled instances
- Add
- Newly labelled

Cooperative Learning and its Application to Emotion Recognition from Speech,”
Cooperative Learning.

- **Example: ComParE:EC**
  1) Active Learning (AL)
  2) Semi-Supervised
  3) Cooperative Learning
     - cross-view (xv)
     - multi-view (mv)
  4) Dynamic
     - further 79.2% reduction

"Cooperative Learning and its Application to Emotion Recognition from Speech,"
iHEARu-PLAY

• **Playful Sourcing**

  *iHEARu-PLAY:*
  Gamified annotation
  Competing w/ others
  Gratification...

<table>
<thead>
<tr>
<th>Badge Name</th>
<th>Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Bird</td>
<td>Answer 100 questions between 4 am and 7 am</td>
</tr>
<tr>
<td>Night Owl</td>
<td>Answer 100 questions between 23 pm and 2 am</td>
</tr>
<tr>
<td>Expert</td>
<td>Reach a score of 5000 Points</td>
</tr>
<tr>
<td>Master</td>
<td>Reach a score of 20000 Points</td>
</tr>
<tr>
<td>Powerman</td>
<td>Collect 100 Bonus Items (in total)</td>
</tr>
<tr>
<td>Regular Customer</td>
<td>Have a constant log-in streak of 7 days in a row</td>
</tr>
<tr>
<td>Way to go</td>
<td>Answer 100 questions in total</td>
</tr>
<tr>
<td>Autobiographer</td>
<td>Fill out own bibliography</td>
</tr>
<tr>
<td>Chatterbox (hidden)</td>
<td>Used the contact form 5 times</td>
</tr>
</tbody>
</table>

Cooperative Learning.

- **Confidence Measure**

  Learning Recogniser Behaviour

  Adaptation to target domain:
  Semi-supervised learning

Distribution.

Embedding.

- **Mobile Platform**
  Samsung Galaxy S3 (Android) vs Intel Core i3 2.1 GHz Laptop

<table>
<thead>
<tr>
<th>Set</th>
<th># feats.</th>
<th>RTF (S3)</th>
<th>RTF (i3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GeMAPS Extended</td>
<td>88</td>
<td>2.63</td>
<td>0.11</td>
</tr>
<tr>
<td>Interspeech Emotion Ch.</td>
<td>384</td>
<td>0.43</td>
<td>0.04</td>
</tr>
<tr>
<td>Interspeech ComParE</td>
<td>6373</td>
<td>2.81</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Learning Hidden Representations.

• Semi-NMF

\[ X^{\pm} \approx Z^{\pm} H^{+} \]

Matrix factorization, uncovers meaningful features

Clustering interpretation:
- Z: cluster centroids
- H: soft membership for every data point \( h_{ik} = \begin{cases} 1 & \text{if } x_i \text{ belongs to cluster} \\ 0 & \text{otherwise} \end{cases} \)

Soft version of k-means clustering – equivalent: iff H is orthogonal with:

\[
C'_{k-\text{means}} = \sum_{i=1}^{n} \sum_{j=1}^{k} h_{ki} \| x_i - z_k \|^2 = \| X - Z H \|^2_F
\]

Better, if data is not distributed in spherical manner
Deep Semi-NMF.

Deep Semi-NMF.

The CMU Multi Pose, Illumination, and Expression (free) 2,856 samples

The CMU Multi Pose, Illumination, and Expression (proprietary) 13,230 samples

The Extended Multi Modal Verification for Teleservices and Security applications 2,360 samples

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Clustering Accuracy</th>
<th>Number of Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deep Semi-NMF</td>
<td>0.65</td>
<td>40</td>
</tr>
<tr>
<td>Semi-NMF</td>
<td>0.60</td>
<td>50</td>
</tr>
<tr>
<td>NMF (MUL)</td>
<td>0.55</td>
<td>60</td>
</tr>
<tr>
<td>Multi-layer NMF</td>
<td>0.50</td>
<td>70</td>
</tr>
<tr>
<td>NeNMF</td>
<td>0.45</td>
<td>80</td>
</tr>
<tr>
<td>GNMF</td>
<td>0.40</td>
<td>90</td>
</tr>
</tbody>
</table>

UAAUC [%] | Pose | Emotion | ID |
----------|------|---------|----|
Semi-NMF  | 94.32| 44.72   | 50.33|
Deep Semi-NMF | 99.78| 73.33   | 74.79|
A priori DSNMF | 99.99| 76.67   | 80.44|

Challenge Your Peers
The Challenge

• **Why at all?**
  
  Provide unified test-beds for evaluation
  Exchange of ideas across the community
  …on specific given tasks

  Help overcome often present lack of comparability
  …due to different data-sets, partitioning, evaluation measures, etc.

  May serve as long-standing reference

  Raises awareness for your field outside of it

  Raise acceptance for “low numbers” (the reality shock)
Risks

• **Focus on Optimisation**

  Novel paradigms often take development time
  Other aspects such as efficiency often neglected
  Awarding best paper can ease focus on optimisation

• **Over-Optimisation**

  Using new datasets for the challenge helps
## Selecting a Corpus

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Example</th>
</tr>
</thead>
</table>
| **Quantity** | “There’s no data like more data”  
High diversity with respect to diverse factors  
Reasonably balanced distribution  
Knowledge of natural distribution (‘priors’) |
| **Quality** | Adequate data  
Realistic data  
Ideal capture conditions  
Intended corruption |
| **Modelling** | Reasonable categorisation  
Well-defined mappings between models |
| **Labelling** | Unique and additional labelling (text+events, labeller tracks, context, ...)  
Provision of gold standard’s reliability |
| **Release** | Documentation of side conditions  
Additional perception tests  
Freeing of the data  
Defined partitioning |
The End User License Agreement

- EULA should
  - Be the same for all corpora
  - Fit one page (fax!)
  - Be signed by permanent staff member
  - Emailed or faxed...
The Data

- Mandatory Things

  Permissions to distribute the data

  An end-user license agreement

  There is a development and a test set

  The test set is not known to anybody previously (!)

  Train and test are sufficiently different…

  It contains the task and not rather noise
Finding a Venue

- To satellite or not to satellite

ACM ICMI, ACM MM, HUMAINE / IEEE ACII, IEEE FG, Interspeech
IEEE ICASSP (?), NIPS (?), ICDAR (?), ICPR (?), IJCAI (?), etc.

Satellite often less effort and more attractive

Potential to reach audience of main conference

More restrictive, e.g., with proceedings
Rules Rule!

- Rules...

  Making rules clear from beginning
  Readable at any time to all (!) participants
  No change during the ongoing Challenge
  Sticking strictly (!) to them
  Clarifying immediately

  Organisers do not participate
  Clarify if (independent) paper acceptance is mandatory
  Limit number of trials
  Preserve right to re-evaluate winners’ results
The Task Design

• Make the Task

  As close to the real thing as possible

  Appealing, novel in some aspect

  Useful to the world

  Not too hard…

  Manageable in terms of data transfer

  Comprehensible
Testing

• **Measures**
  Accuracy (Un-/Weighted Average Recall)
  Precision, F-Measure, confusions
  Area under ROC
  Mean Linear Error, Cross Correlation

• **Partitioning**
  J-fold Stratified Cross-Validation
  Leave-One-Speaker (group)-Out
  Train-Develop-Test

• **Discussion**
  Significance
  Ground truth
  Prototypicality
Tools

• **Tasks:**
  
  - Feature extraction
  - Model training and classification
  - Pre-trained models

• **Application Domains:**
  
  - Batch evaluations / Off-line analysis
  - On-line / realtime recognition
The Baseline

• **Has to be:**
  
  State-of-the-art  
  Transparent („no“ optimisation)  
  Reproducible  
  Comprehensible

• **Should be:**
  
  Established by open-source tools  
  Re-doable “at no effort” (e.g., batch scripts provided)  
  Not overly optimised  
  Beatable…
The Paper

- **The Essential**
  
  Describe it such that the participants know

- **The Secret Ingredient**
  
  Highlight one technical aspect to make it more interesting:
  
  IS 2009: Instance balancing / dynamic vs. static
  IS 2010: Multitask vs. single task / regression task
  IS 2011: Feature optimisation
  IS 2012: Effect of random seed (standard deviation)
  IS 2013: Detection
Evaluation

• **Saving work…**
  
  Have a web script for upload
  
  Not requiring codes and running them, but
  Keep right to ask for it

• **Ensure that**
  
  The servers is up
  
  You provide a “safe” deadline
The Workshop / Event / Session

• **Instruct the Participants**
  Write mail in advance
  Explain presentation format
  Potentially limit (e.g., no repetition of challenge conditions, data, etc.)

• **Choose format wisely**
  Poster session usually not suited – participants are curious to see others
  Short talks and intro / summary good choice

• **Keep the suspension**
  Consider announcing winners at end
  Lay out presentations in suited order
Awarding

• Make a Certificate

• Make a Dummy Cheque

INTER SPEECH 2012 Speaker Trait Challenge

<table>
<thead>
<tr>
<th>Award Cheque</th>
</tr>
</thead>
<tbody>
<tr>
<td>[EUR]</td>
</tr>
<tr>
<td>€</td>
</tr>
</tbody>
</table>

Pay to:

The order of (in words):

Drawer: 

Authorized Signature:

Monday, September 10, 2012

Portland, Oregon, U.S.A.

Organizers of the Speaker Trait Challenge

• Award in the big Ceremony

• Put the Winners on the Web
Evaluation

- Fusion

  Late fusion (vote)

  Ask for confidences

  Determining optimal N

  Usually long-standing

  Upper benchmark

  Message: “Together we’re best”
Evaluation

- **Significance**

  Heavily depends on:

  # test instances
  % baseline

## Evaluation

- **Q-Statistics**

  Pairwise measuring whether participants commit same errors on test

<table>
<thead>
<tr>
<th></th>
<th>Avg</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
<th>#6</th>
<th>#7</th>
<th>#8</th>
<th>#9 (Nogueiras)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1 (Bone et al.)</td>
<td>.427</td>
<td>.598</td>
<td>.425</td>
<td>.461</td>
<td>.460</td>
<td>.428</td>
<td>.433</td>
<td>.395</td>
<td>.217</td>
</tr>
<tr>
<td>#2 (Bocklet et al.)</td>
<td>.657</td>
<td>.710</td>
<td>.752</td>
<td>.757</td>
<td>.690</td>
<td>.740</td>
<td>.651</td>
<td>.356</td>
<td></td>
</tr>
<tr>
<td>#3 (Höning et al.)</td>
<td>.754</td>
<td>.822</td>
<td>.941</td>
<td>.906</td>
<td>.946</td>
<td>.864</td>
<td>.418</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#4 (Gajsek et al.)</td>
<td>.716</td>
<td>.836</td>
<td>.845</td>
<td>.880</td>
<td>.775</td>
<td>.358</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#5 (Huang et al.)</td>
<td>.776</td>
<td>.917</td>
<td>.956</td>
<td>.885</td>
<td>.452</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#6 (Ultes et al.)</td>
<td>.749</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#7 (Baseline)</td>
<td>.782</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#8 (Montacié and Caraty)</td>
<td>.706</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#9 (Nogueiras)</td>
<td>.374</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continuity

• To turn it into a continued success...

  Make it consistent

  Use same EULA, format, conditions, etc.

  Alter … depending on last experiences
The Repository

- **Easy access to data**

  Make a page with info on past challenges

  Provide info for data access (contact persons, EULAs, etc.)

  Provide baseline feature sets (the ML community will appreciate it!)

  Make it a value way beyond the singular event

  Save yourself 1000 emails…

  And … really do it! (coming up some day soon for ComParE / AVEC…)
Wrap-Up

- **Organise a Challenge**
  
  Find great data  
  Set clear rules  
  Provide transparent redoable baselines  
  Provide good infrastructure for data access, result upload, etc.

- **In the long run**
  
  Make a Special Issue  
  Ensure consistency  
  Make a repository  
  Participate in other challenges
Experience ComParE / AVEC

• **# Participating Teams**


• **Some Observations**

  Fusion of best participant results exceeds individual winners throughout
  These “multiple-site” results are so far not reached by individual attempts
  Supra-segmental modelling of information prevails by far
  Only sparsely dynamic algorithms such as HMMs or DBNs
  Cross disciplines seems to be successful
  Great submissions outside competition, e.g., perception studies
Outlook
And Next...

- Big Data Multi-task DL
- 24/7 Coop Learning
- “Green” Learning
- Reinforced Learning
- Evolving Machines
- New Challenges
- Products!
Better Ideas?

- Special Issue on New Avenues in Knowledge Bases for Natural Language Processing of Knowledge-Based Systems, Elsevier
- "Emotion Representations and Modelling for Companion Technologies" (ERM4CT) workshop 2015 @ 17th ACM ICMI 2015, Seattle, WA, 09.-13.11.2015.
- AV+EC 2015, @ ACM Multimedia, Brisbane, Australia, 26.-30.10.2015.
- 1st International Workshop on Automatic Sentiment Analysis in the Wild (WASA 2015) @ ACII 2015, Xian, PR China, 21.09.2015.
- Special Session Socio-cognitive Language Processing @ 7th International Workshop on Spoken Dialogue Systems (IWSDS), Riekonlinna 2016.