Implicit Feedback for User-Adaptive Systems by Analyzing the Users’ Speech

Frank Wittig / Christian Müller
Saarland University, Germany
{wittig, cmueller}@cs.uni-sb.de

Outline

- Motivation: Universal Usability
- M3I
- Speech as a Source for UM
- Two-Layered ML Recognition Approach
- Summary / Current and Future Work
Universal Usability

context diversity  (technical resources)
- large displays
- fast CPU
- large memory
- ... 
- small displays
- slow CPU
- small memory
- ...

user diversity  (cognitive resources)
- high WM load
- time pressure
- good eyesight
- ...
- low WM load
- no time pressure
- poor eyesight
- ...

M3I
- a Mobile Multi-modal and Modular Interface
- developing a framework for resource-adaptive mobile multi-modal systems
  - recognizing both technical and cognitive resource limitations
  - adapting the systems behaviour accordingly
- applications
  - mobile pedestrian navigation system
  - shopping assistant
  - museum guide
This talk

- a Mobile Multi-modal and Modular Interface
- developing a framework for resource-adaptive mobile multi-modal interfaces
  - recognizing both technical and cognitive resource limitations
  - adapting the systems behaviour accordingly
- applications
  - mobile pedestrian navigation system
  - shopping assistant
  - museum guide

Simplification

- The notion of cognitive resources is restricted to the distinction between

  „average aged adults“ ← the elderly
Speech as a Source for UM

Linguistic Features (UM01)
i.e. self-corrections, repetitions

Prosodic Features
i.e. articulation rate

Acoustic Features
i.e. jitter, shimmer

expensive
NLP necessary
Sensitive to changes in the acoustic environment

Classification Tree

speaker

Jitter
Shimmer

male

female

elderly
non-elderly
elderly
non-elderly

"voices of men and women age differently"
Two-Layered Machine Learning Approach

- **Classification Layer**
  - artificial neural network
  - classifies a given speaker (uncertain)

- **Meta-Reasoning Layer**
  - dynamic Bayesian network
  - treats uncertainty
  - models gender specific aging
  - improves model incrementally

---

**Classification Layer**

- **Corpus**

<table>
<thead>
<tr>
<th>Number of speakers</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>male</td>
<td>586</td>
</tr>
<tr>
<td>female</td>
<td>442</td>
</tr>
<tr>
<td>elderly</td>
<td>676</td>
</tr>
<tr>
<td>non-elderly</td>
<td>352</td>
</tr>
</tbody>
</table>
Classification Layer

Corpus $\rightarrow$ Feature Extraction:
- Jitter
- Shimmer
(11 features in total)

Classification Layer

Corpus $\rightarrow$ Feature Extraction $\rightarrow$ Artificial Neural Network

- performed best in initial study compared with
  C 4.5 decision tree induction, k-nearest neighbors,
  naive Bayes, and support vector machines
### Alternative Classifiers

Prediction Accuracy (10-fold CV, WEKA):

<table>
<thead>
<tr>
<th></th>
<th>C4.5</th>
<th>ANN</th>
<th>kNN</th>
<th>NB</th>
<th>SVM</th>
<th>BL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>88.07</td>
<td>93.14</td>
<td>91.55</td>
<td>81.64</td>
<td>86.07</td>
<td>57.84</td>
</tr>
<tr>
<td>Age</td>
<td>85.30</td>
<td>92.88</td>
<td>88.10</td>
<td>70.27</td>
<td>79.26</td>
<td>64.16</td>
</tr>
</tbody>
</table>

### Classification Layer

- Corpus
- Feature Extraction
- Artificial Neural Network
- Classifier
**Classification Layer**

Classification accuracy (10-fold cross validation)

- Male: ~96%
- Female: ~89%
- Elderly: ~88%
- Non-elderly: ~96%

**Meta-Reasoning Layer**

- Gender classifier: male 0.86, female 0.14
- Age classifier:
  - Male: 1.0
  - Female: 0.0
  - Non-elderly: 0.85
  - Elderly: 0.15
**Meta-Reasoning Layer**

- Corpus
- Feature Extraction
- Artificial Neural Network
- (Specialized) Classifiers

- CPTs: True positive rates
- Cross-validation
- Training-/Test-Sets

**Learn BN (CPTs)**

**Meta-Reasoning Layer (future extensions)**

- Gender
- Age
- Dialog history
- Gender classifier
- Gender cl. (pitch)
- Age male classifier
- Age female classifier
- Biometric sensors
- Context
  - Microphone quality
  - Background noise
Continuously Updating the Model

Summary and Future Work

- two-layered machine-learning approach for estimating age and gender of a speaker
  - classification layer (ANN):
    - high accuracy using two acoustic measures jitter and shimmer.
  - meta-reasoning layer (DBN):
    - modeling uncertainty and gender-specific aging
    - incrementally updating the user model
- the approach is implemented in a client/server architecture
- main issues of (near) future work
  - integration into testbed applications
    - pedestrian navigation and shopping/museum guide
  - developing adaptation strategies

Thank you very much for your attention!