Feature Selection for Speaker Traits

Jouni Pohjalainen\textsuperscript{1}, Serdar Kadioglu\textsuperscript{2}, Okko Räsänen\textsuperscript{1}

\textsuperscript{1} Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland, jouni.pohjalainen@aalto.fi, okko.rasanen@aalto.fi

\textsuperscript{2} Department of Computer Science, Brown University, Providence, RI 02912, USA, serdark@cs.brown.edu
Overview

- Introduction
- Feature selection methods
- Classification method
- Evaluation procedure
- Likability sub-challenge
- Pathology sub-challenge
- Personality sub-challenge
- Conclusions
Introduction

- 6125 utterance-level features were provided or each sub-challenge

- Instead of focusing on classification methods for high-dimensional data, can we find feature spaces where classification is “easy”?

- There are are $2^{6125} - 1$ possible feature spaces
Introduction

- Compared to state-of-the-art classification methods for high-dimensional data, can we reach their performance by means of just feature selection and a basic classification method?

- The results should be generalizable
  - avoid overlearning any single feature selection objective
## Approach 1 - Classification using individual features

<table>
<thead>
<tr>
<th>Audio Clip 1</th>
<th>Feature 1</th>
<th>Feature 2</th>
<th>...</th>
<th>Feature m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio Clip 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audio Clip n</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- We start with a classification phase
Approach 1 - Classification using individual features

<table>
<thead>
<tr>
<th>Audio Clip 1</th>
<th>Feature 1</th>
<th>Feature 2</th>
<th>...</th>
<th>Feature m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio Clip 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audio Clip n</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- We start with a classification phase
- **Goal**: classify the speech audio clips using each individual feature
### Approach 1 - Classification using individual features

- **Supervised training**
  - Train 8-component GMMs for both classes
  - Trait is present or not

<table>
<thead>
<tr>
<th>Audio Clip 1</th>
<th>Feature 1</th>
<th>Feature 2</th>
<th>...</th>
<th>Feature m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio Clip 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audio Clip n</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Approach 1 - Classification using individual features

<table>
<thead>
<tr>
<th>Feature 1</th>
<th>Feature 2</th>
<th>...</th>
<th>Feature m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio Clip 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audio Clip 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audio Clip n</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Unsupervised training**

  - After initial supervised training, combine the two GMMs into one and let it freely adapt to the training data during additional EM iterations, then again separate the two GMMs
Approach 1 - Classification using individual features

<table>
<thead>
<tr>
<th>Feature 1</th>
<th>Feature 2</th>
<th>…</th>
<th>Feature m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio Clip 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audio Clip 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>…</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audio Clip n</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Use the GMMs to classify the clips according to the **Bayes** rule
- Classify the development set using GMMs trained on the training set and vice versa
Approach 1 - Classification using individual features

<table>
<thead>
<tr>
<th>Feature 1</th>
<th>Feature 2</th>
<th>…</th>
<th>Feature m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio Clip 1</td>
<td>0 / 1</td>
<td>…</td>
<td></td>
</tr>
<tr>
<td>Audio Clip 2</td>
<td></td>
<td>…</td>
<td></td>
</tr>
<tr>
<td>…</td>
<td></td>
<td>…</td>
<td></td>
</tr>
<tr>
<td>Audio Clip n</td>
<td></td>
<td></td>
<td>0 / 1</td>
</tr>
</tbody>
</table>

- for each audio clip
- for each feature
  - note whether the classification was **correct** or **not**
## Feature Selection

- **Properties we want**
  - For each clip there must be at least one correct classification
  - Number of selected features minimized

<table>
<thead>
<tr>
<th></th>
<th>Feature 1</th>
<th>Feature 2</th>
<th>...</th>
<th>Feature m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio Clip 1</td>
<td>0 / 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audio Clip 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td>...</td>
</tr>
<tr>
<td>Audio Clip n</td>
<td></td>
<td></td>
<td></td>
<td>0 / 1</td>
</tr>
</tbody>
</table>
Observation

We can formulate this as the **Set-Covering Problem**

- Features as *sets*
- Audio clips as *items*
Observation

<table>
<thead>
<tr>
<th>Feature 1</th>
<th>Feature 2</th>
<th>...</th>
<th>Feature m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio Clip 1</td>
<td>0 / 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audio Clip 2</td>
<td></td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audio Clip n</td>
<td></td>
<td></td>
<td>0 / 1</td>
</tr>
</tbody>
</table>

- We can formulate this as the Set-Covering Problem
  - Features as sets
  - Audio clips as items
- Each set has a cost of one: \textit{unicost-SCP}
Set-Covering Problem (SCP)

<table>
<thead>
<tr>
<th>Feature 1</th>
<th>Feature 2</th>
<th>...</th>
<th>Feature m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio Clip 1</td>
<td>0 / 1</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Audio Clip 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audio Clip n</td>
<td></td>
<td></td>
<td>0 / 1</td>
</tr>
</tbody>
</table>

- **Goal:**
  - Select features/sets/columns such that each clip/item/row is covered at least once
  - Total cost (number of selected features) is minimized
Solving the SCP

- This observation has an immediate bearing on the feature selection problem

- We can leverage general techniques for solving the SCP
Solving the SCP

- **Exact solution method**
  - Branch-and-bound algorithm using the linear relaxation
Solving the SCP

- Exact solution method
  - Branch-and-bound algorithm using the linear relaxation
- In general, SCP is NP-hard
Solving the SCP

- **Approximate solution methods**
  - **Greedy algorithm**: select iteratively the features that cover the most remaining clips
  - **Rounding-up technique**: solve the linear relaxation. Then, to get an integer solution, round-up the values of every fractional value to 1 (i.e. select these sets)
Solve the SCP

<table>
<thead>
<tr>
<th>Feature 1</th>
<th>Feature 2</th>
<th>…</th>
<th>Feature m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio Clip 1</td>
<td>0 / 1</td>
<td>…</td>
<td>0 / 1</td>
</tr>
<tr>
<td>Audio Clip 2</td>
<td>0 / 1</td>
<td>…</td>
<td>0 / 1</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>Audio Clip n</td>
<td>0 / 1</td>
<td>…</td>
<td>0 / 1</td>
</tr>
</tbody>
</table>

Based on supervised classification: SSCP
Solve the SCP

<table>
<thead>
<tr>
<th></th>
<th>Feature 1</th>
<th>Feature 2</th>
<th>...</th>
<th>Feature m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio Clip 1</td>
<td>0 / 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audio Clip 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audio Clip n</td>
<td></td>
<td></td>
<td></td>
<td>0 / 1</td>
</tr>
</tbody>
</table>

- Based on supervised classification: SSCP
- Based on unsupervised classification: USCP
Approach II – Statistical dependence between features and labels

- Discretize each feature by quantizing it to one of \( N = 65 \) levels, where the quantization scale is adjusted s.t. each bin will contain an equal amount of samples.

- Measure the statistical dependence (SD) using the formula

\[
D = \sum_{y \in Y} \sum_{z \in Z} p(y, z) \frac{p(y, z)}{p(y)p(z)}
\]

where \( y \) is the discretized feature and \( z \) is the class labeling.
Approach II – Statistical dependence between features and labels

- For comparison, the conventional mutual information (MI) measure is given by

\[ D = \sum_{y \in Y} \sum_{z \in Z} p(y, z) \log \left( \frac{p(y, z)}{p(y)p(z)} \right) \]

- In order to determine the number of features to be chosen according to the SD measure in the combined selection methods, subsets of each size of the SCP-based feature sets were obtained by random selection and by SD (using training data)
Combined feature selection methods

- First select a few hundred features using SSCP or USCP

- Use the SD or MI measure to select features among this subset
Classification method: \( k \) nearest neighbors (kNN)

- Each feature is normalized to have zero mean and unit variance
- Find \( k \) nearest neighbors according to the Euclidean distance measure
- Among the \( k \) nearest neighbors, the counts of different classes are scaled by dividing them by the frequencies of occurrence of those classes in the training data
- Value of \( k \) determined experimentally
## Likability sub-challenge

<table>
<thead>
<tr>
<th>Data</th>
<th>Method of feature selection</th>
<th>Number of features</th>
<th>k</th>
<th>UA</th>
<th>WA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Devel</td>
<td>Baseline (SVM)</td>
<td>6125</td>
<td></td>
<td>58.5</td>
<td></td>
</tr>
<tr>
<td>Devel</td>
<td>SSCP</td>
<td>406</td>
<td>60</td>
<td>57.4</td>
<td></td>
</tr>
<tr>
<td>Devel</td>
<td>USCP</td>
<td>423</td>
<td>35</td>
<td>59.8</td>
<td></td>
</tr>
<tr>
<td>Devel</td>
<td>SD</td>
<td>349</td>
<td>65</td>
<td>55.3</td>
<td></td>
</tr>
<tr>
<td>Devel</td>
<td>SSCP+SD</td>
<td>349</td>
<td>90</td>
<td>62.0</td>
<td></td>
</tr>
<tr>
<td>Devel</td>
<td>USCP+SD</td>
<td>411</td>
<td>40</td>
<td>60.0</td>
<td></td>
</tr>
<tr>
<td>Devel</td>
<td>SSCP+MI</td>
<td>349</td>
<td>90</td>
<td>58.6</td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>Baseline (RF)</td>
<td>6125</td>
<td></td>
<td>59.0</td>
<td>59.2</td>
</tr>
<tr>
<td>Test</td>
<td>USCP+SD</td>
<td>411</td>
<td>58</td>
<td>53.3</td>
<td>53.9</td>
</tr>
<tr>
<td>Test</td>
<td>SSCP+SD</td>
<td>349</td>
<td>131</td>
<td>61.3</td>
<td>61.3</td>
</tr>
</tbody>
</table>
## Pathology sub-challenge

<table>
<thead>
<tr>
<th>Data</th>
<th>Method of feature selection</th>
<th>Number of features</th>
<th>k</th>
<th>UA</th>
<th>WA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Devel</td>
<td>Baseline (RF)</td>
<td>6125</td>
<td></td>
<td>64.8</td>
<td></td>
</tr>
<tr>
<td>Devel</td>
<td>SSCP</td>
<td>318</td>
<td>115</td>
<td>67.2</td>
<td></td>
</tr>
<tr>
<td>Devel</td>
<td>USCP</td>
<td>347</td>
<td>85</td>
<td>67.2</td>
<td></td>
</tr>
<tr>
<td>Devel</td>
<td>SD</td>
<td>300</td>
<td>30</td>
<td>66.2</td>
<td></td>
</tr>
<tr>
<td>Devel</td>
<td>SSCP+SD</td>
<td>299</td>
<td>75</td>
<td>68.3</td>
<td></td>
</tr>
<tr>
<td>Devel</td>
<td>USCP+SD</td>
<td>172</td>
<td>125</td>
<td>63.8</td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>Baseline (RF)</td>
<td>6125</td>
<td></td>
<td>68.9</td>
<td>67.5</td>
</tr>
<tr>
<td>Test</td>
<td>USCP</td>
<td>347</td>
<td>155</td>
<td>65.5</td>
<td>72.4</td>
</tr>
<tr>
<td>Test</td>
<td>USCP+SD</td>
<td>300</td>
<td>137</td>
<td>65.6</td>
<td>72.4</td>
</tr>
<tr>
<td>Test</td>
<td>SSCP+SD</td>
<td>299</td>
<td>137</td>
<td>66.3</td>
<td>69.8</td>
</tr>
</tbody>
</table>
# Personality sub-challenge

<table>
<thead>
<tr>
<th>Data</th>
<th>Method of feature selection</th>
<th>Number of features</th>
<th>k</th>
<th>UA</th>
<th>WA</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>Devel / Devel</td>
<td>Baseline (SVM)</td>
<td>6125</td>
<td>50</td>
<td>60.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SSCP+SD</td>
<td>277</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>Test</td>
<td>Baseline (RF)</td>
<td>6125</td>
<td>86</td>
<td>59.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SSCP+SD</td>
<td>277</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Devel / Devel</td>
<td>Baseline (RF)</td>
<td>6125</td>
<td>40</td>
<td>74.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>USCP+SD</td>
<td>285</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>Test</td>
<td>Baseline (SVM)</td>
<td>6125</td>
<td>70</td>
<td>80.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>USCP+SD</td>
<td>176</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>USCP+SD</td>
<td>285</td>
<td>69</td>
<td>79.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>76.8</td>
</tr>
</tbody>
</table>
## Personality sub-challenge

<table>
<thead>
<tr>
<th>Data</th>
<th>Method of feature selection</th>
<th>Number of features</th>
<th>k</th>
<th>UA</th>
<th>WA</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>Baseline (RF) USCP+SD</td>
<td>6125</td>
<td>70</td>
<td>82.8</td>
<td>85.8</td>
</tr>
<tr>
<td></td>
<td>Devel</td>
<td>291</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Test</td>
<td>6125</td>
<td>120</td>
<td>76.2</td>
<td>72.6</td>
</tr>
<tr>
<td></td>
<td>Test</td>
<td>291</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Baseline (SVM) USCP+SD</td>
<td>6125</td>
<td>15</td>
<td>67.6</td>
<td>72.1</td>
</tr>
<tr>
<td>A</td>
<td>Devel SSPCP+SD</td>
<td>217</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Test</td>
<td>6125</td>
<td>26</td>
<td>64.2</td>
<td>64.2</td>
</tr>
<tr>
<td></td>
<td>Test</td>
<td>217</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>Baseline (RF) SSPCP+SD</td>
<td>6125</td>
<td>40</td>
<td>68.9</td>
<td>74.8</td>
</tr>
<tr>
<td></td>
<td>Devel SSPCP+SD</td>
<td>124</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Test</td>
<td>6125</td>
<td>69</td>
<td>65.9</td>
<td>67.6</td>
</tr>
<tr>
<td></td>
<td>Test</td>
<td>124</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusions

Using feature selection and simple kNN classification, it was possible to achieve classification accuracy comparable to state-of-the-art classification methods using the full high-dimensional feature space.

Good results were obtained by two approaches:

- Set covering problem (SCP) based on classification using single features (different classification method)
- Statistical dependence (SD) between discretized features and class labeling
Conclusions

- The results generalized from one dataset to another

- By combining differently based feature selection criteria (SCP and SD), better results were obtained than by using a single criterion