Features selected by hill-climbing with five streams

December 2, 2008

1 Introduction

This document contains diagrams showing what features were chosen by hill-climbing with the Numbers corpus using five streams, along with the initial feature vectors that hill-climbing started from. The full writeup about these experiments is in an ICSI technical report by David Gelbart. The technical report should be finished by the end of December 2008 and the currently planned title is “Hill-climbing ensemble feature selection with more streams”. It is a followup to David Gelbart’s PhD thesis.

The five streams were the 13 static MFCC features, the 26 dynamic (delta and delta-delta) MFCC features, the 13 static PLP features, the 26 dynamic PLP features, and the 28 MSG features.

In each diagram, the boxes are divided into rows. The first row is MFCC static features (the first 13 MFCC features), the second row is MFCC delta features (the next 13 MFCC features), the third row is MFCC delta-delta features, then PLP static features, PLP delta features, PLP delta-delta features, and finally MSG features. The shaded boxes correspond to features that are included in the feature vector, and the empty boxes correspond to features that are not included.
## 2 Initial feature vectors

<table>
<thead>
<tr>
<th>Feature</th>
<th>Stream 1</th>
<th>Stream 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFCC</td>
<td><img src="stream1_mfcc.png" alt="Feature Vector" /></td>
<td><img src="stream1_mfcc.png" alt="Feature Vector" /></td>
</tr>
<tr>
<td>MFCC Δ</td>
<td><img src="stream1_mfcc_delta.png" alt="Feature Vector" /></td>
<td><img src="stream1_mfcc_delta.png" alt="Feature Vector" /></td>
</tr>
<tr>
<td>MFCC ΔΔ</td>
<td><img src="stream1_mfcc_deltadelta.png" alt="Feature Vector" /></td>
<td><img src="stream1_mfcc_deltadelta.png" alt="Feature Vector" /></td>
</tr>
<tr>
<td>PLP</td>
<td><img src="stream1_plp.png" alt="Feature Vector" /></td>
<td><img src="stream1_plp.png" alt="Feature Vector" /></td>
</tr>
<tr>
<td>PLP Δ</td>
<td><img src="stream1_plp_delta.png" alt="Feature Vector" /></td>
<td><img src="stream1_plp_delta.png" alt="Feature Vector" /></td>
</tr>
<tr>
<td>PLP ΔΔ</td>
<td><img src="stream1_plp_deltadelta.png" alt="Feature Vector" /></td>
<td><img src="stream1_plp_deltadelta.png" alt="Feature Vector" /></td>
</tr>
<tr>
<td>MSG</td>
<td><img src="stream1_msg.png" alt="Feature Vector" /></td>
<td><img src="stream1_msg.png" alt="Feature Vector" /></td>
</tr>
</tbody>
</table>

Figure 1: Initial feature vector for stream 1.

Figure 2: Initial feature vector for stream 2.
Figure 3: Initial feature vector for stream 3.

Figure 4: Initial feature vector for stream 4.

Figure 5: Initial feature vector for stream 5.
3 Final feature vectors for the clean Numbers corpus. Equation 5.1 from the PhD thesis (Opitz’s formula) was used to guide hill-climbing.

![Figure 6](image1.png)

**Figure 6:** Final feature vector for stream 1.

![Figure 7](image2.png)

**Figure 7:** Final feature vector for stream 2.
Figure 8: Final feature vector for stream 3.

Figure 9: Final feature vector for stream 4.

Figure 10: Final feature vector for stream 5.
4 Final feature vectors for the noisy Numbers corpus. Equation 5.1 was used to guide hill-climbing.

Figure 11: Final feature vector for stream 1.

Figure 12: Final feature vector for stream 2.
Figure 13: Final feature vector for stream 3.

Figure 14: Final feature vector for stream 4.

Figure 15: Final feature vector for stream 5.