Convolutional Deep Maxout Networks for Phone Recognition

László Tóth

Abstract

Convolutional neural networks (CNN) can achieve excellent speech recognition results, thanks to the pooling operation used to aggregate neural activations. In this work, we investigated the possibility of replacing the pooling operation with maxout units. The results show that maxout networks perform better or equally as well as ReLU networks.

2 Conclusion

Maxout units evaluate the different inputs of a neuron and perform the maximization in one step. This operation can be seen as a combination of maxout and 2-norm pooling.

3 Experiments on TIMIT

The input features were 123 FBANK values, both on the development set and the test set. First, the linear activation of each neuron is calculated. Then, the maxout activation function divides the maximum of the linear neural activations in the given group.

4 Introduction

Convolutional maxout neurons execute both types of evaluation and perform the maximization in one step.

5 Results with Convolutional Deep Maxout Networks

<table>
<thead>
<tr>
<th>Network type</th>
<th>Phone Error Rate</th>
<th>Frame Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>ReLU</td>
<td>17.0%</td>
<td>13.9%</td>
</tr>
<tr>
<td>Maxout</td>
<td>19.3%</td>
<td>17.5%</td>
</tr>
</tbody>
</table>

The maxout network performed better for all network configurations investigated, and yields relative error rate reductions of between 2% and 6%.