Ngan, Ganapathiraju, and Picone attempt to overcome the difficult problem of proper noun pronunciation. In this paper, the researchers approach the problem differently than previous attempts. They present a decision tree method approach, and contrast this with a Boltzmann approach (a neural network technique) for speech recognition.

Designing a method for proper noun pronunciations is a difficult task. The problem stems from the fact that the pronunciations do not follow any particular set of rules. Proper nouns such as surnames often have ethnic/cultural origins and often do not follow typical English phoneme pronunciations. The Classical rule based approach generates a single pronunciation and is unsuitable since a given surname can have multiple pronunciations. What is needed is a network that covers all suitable pronunciations that are accepted.

Using a decision tree, the researchers have designed a system that improves previous attempts at proper noun recognition. The decision tree models all possible pronunciations. Decision trees are created using information within the test data. Each node represents a question and each branch splits the distribution of the test data. Beginning with the top node and moving downward, each question is chosen that is most likely to distribute evenly the training data within the tree.

The system generates the phoneme-pronunciation pairs using a training set of name-pronunciation pairs, a context window of fixed length, and n-tuple of letters. For instance, they present the following example for the name “Matt”:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>m</th>
<th>a</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>m</td>
<td>a</td>
<td>t</td>
<td>t</td>
</tr>
<tr>
<td>m</td>
<td>a</td>
<td>t</td>
<td>_</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>t</td>
<td>_</td>
<td>_</td>
<td></td>
</tr>
</tbody>
</table>

At each row in the example, from the context, phoneme pronunciation is generated for
the middle character. Then the characters are shifted to the left and the process is continued. On the rightmost column is the phoneme pronunciation symbol that is generated. Each symbol corresponds to a particular pronunciation.

Once the system has been trained, recognition is performed using a fixed width window. Then for each n-tuple input, the system generates a pronunciation of the surname based upon the context within the window, and the samples it learned from.

An experiment was performed to see how this approach compared to that of the Boltzmann machine (BM) approach. The system was trained using a context of length three. The phoneme error corresponds to the pronunciation of the given phoneme in question, and the name error corresponds to the algorithm correctly pronouncing the entire name. The BM had a phoneme error rate of 37.88% and a name error rate of 70.44%. The decision tree method had a phoneme error rate of 13.28% and a name error rate of 45.5%.

Thus, decision trees were found to be an improved approach to surname pronunciations. The system was written in C++ and is a public domain project.

References