IMPROVED SURNAME PRONUNCIATIONS USING DECISION TREES

Julie Ngan, Aravind Ganapathiraju, Joseph Picone

Institute for Signal and Information Processing
Department of Electrical and Computer Engineering
Mississippi State University,
Mississippi State, MS-39762
{ngan, ganapath, picone}@isip.msstate.edu

SUMMARY

Proper noun pronunciations in speech recognition is a particularly challenging problem since a large percentage of proper nouns often defy typical letter-to-sound conversion rules. There is renewed interest in this problem with the recent decision by the LVCSR community to adopt the named entity task as the next step towards a speech understanding framework for common evaluations. Previous attempts to generate pronunciations from letter context [1] have met with mixed success. A neural network featuring a Boltzmann machine [2] was constructed to generate an ordered list of the N most likely pronunciations of surnames. Various parameters such as context length and number of hidden layers were used to find the best combination for the problem. To train and evaluate the system, we have compiled a hand-transcribed phonetic proper noun database of approximately 18,000 surnames and 25,000 pronunciations. This database adheres to the Worldbet notation for phonetic transcriptions to incorporate data from linguistically diverse sources. It also makes use of dynamic text-to-phoneme alignment [3] (e.g. The surname 'Wright' is transcribed and aligned as '_ 9r aI _ _ t'). However, the best published result on this task was an overall error rate of 52%. Further, the network used to achieve this result also has the disadvantage of being CPU expensive and does not scale up gracefully to handle more complicated tasks.

Statistical decision trees have been used in many disciplines and in various applications for speech recognition, since they provide a flexible and adaptable means for classifying non-linearly separable data. However, currently available decision tree software packages do not handle problems of scale. Moreover, these packages can only handle alphanumeric characters and are unable to handle special characters such as '&', '@', '^', etc. which are used extensively in the proper noun phoneme database. As part of this paper, we will introduce a decision tree software package that is being developed at the Institute for Signal and Information Processing (ISIP) that handles large amounts of data, supports an unlimited number of attributes for classes, and different combinations of splitting, stopping, pruning, and smoothing algorithms. Furthermore, our software also allows data tagging, which enables each attribute to be selected from the attribute file without having to reformat the training data for each experiment. The software will be released shortly to the public domain at *http://www.isip.msstate.edu/*.

In our decision tree system for the proper noun pronunciation problem, a proper noun is padded with null starting and ending symbols. N-tuple of letters of the proper noun are created using a sliding window of a fixed context length. Using this context, the system generates a phoneme pronunciation of each window which are rearranged to generate the pronunciation of the full name. We have evaluated multiple decision tree algorithms on a subset of the database consisting of four-letter names using various combinations of decision tree algorithms including two-ing, Bayesian splitting and smoothing, information gain, gain ratio, etc. The best overall system is a binary, univariate tree that is split using maximum gain ratio and the average information gain per split. This approach produced a phone-based error rate of 13% as opposed to the neural network approach's 22%. We project the corresponding surname error rate to represent a 25% reduction over previous techniques, and will present these results at the conference. This is the first public domain decision tree package that has been successfully applied to such a large speech-related classification task on which other non-linear classifiers have failed.

1 N. Deshmukh and J. Picone, "Automatic Generation of N-Best Pronunciations of Proper Nouns," submitted to the IEEE Transactions on Speech and Audio Processing, November 1996

2 N. Deshmukh, M. Weber, and J. Picone, "Automated Generation of N-Best Pronunciations of Proper Nouns," Proceedings of ICASSP'96, pp. I283-I286, Atlanta, GA, May 1996.

3 N. Deshmukh, J. Ngan, J. Hamaker and J. Picone, "An Advanced System to Generate Multiple Pronunciations of Proper Nouns," Proceedings of ICASSP'97, vol. 2, pp. 1467-1470, Munich, Germany, April 1997.