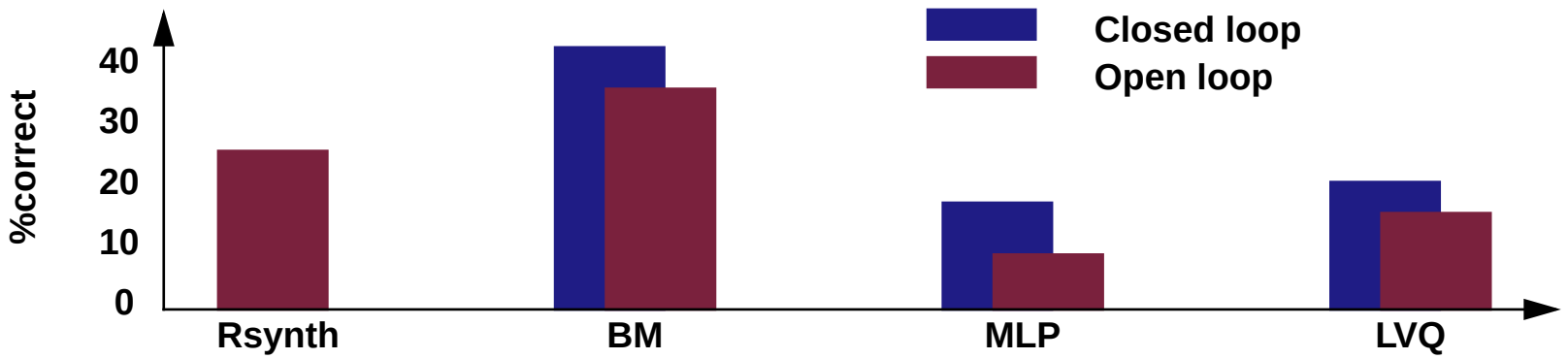
Improved Surname Pronunciation Using Decision Trees

J. Ngan, N. Deshmukh, A. Ganapathiraju, and J. Picone
Institute for Signal and Information Processing
Mississippi State University
{ngan, deshmukh, ganapath, picone}@isip.msstate.edu
http://www.isip.msstate.edu

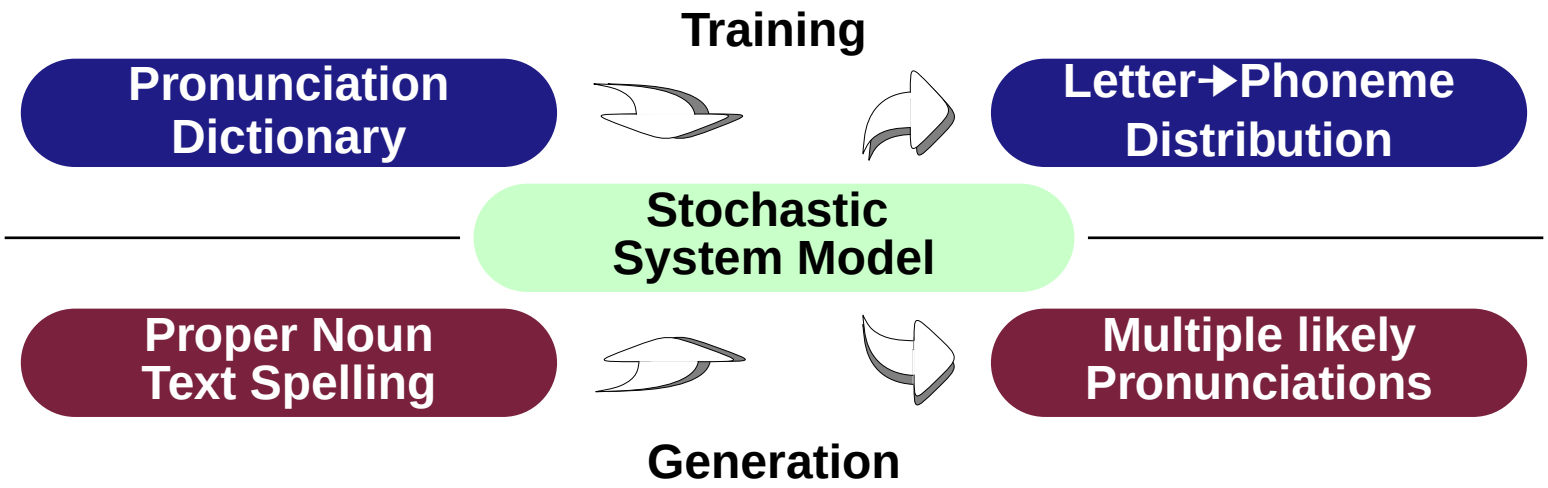
Introduction

- Proper nouns and errors in LVCSR
- Pronunciation networks needed
- Direct letter-to-sound rules do not apply
- Rule-based systems are unsuitable, NN-based systems do not work



Motivation

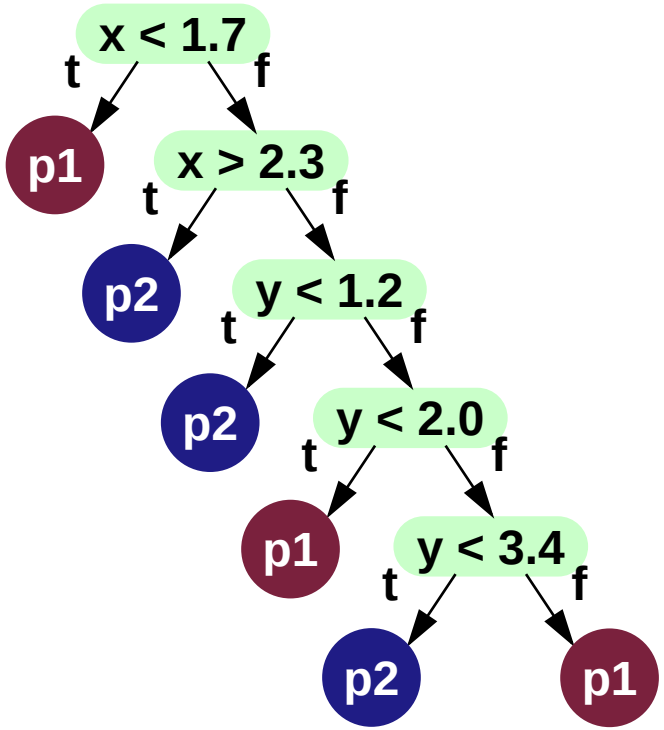
- Classify groups of letters into phonemes using nonlinear stochastic models

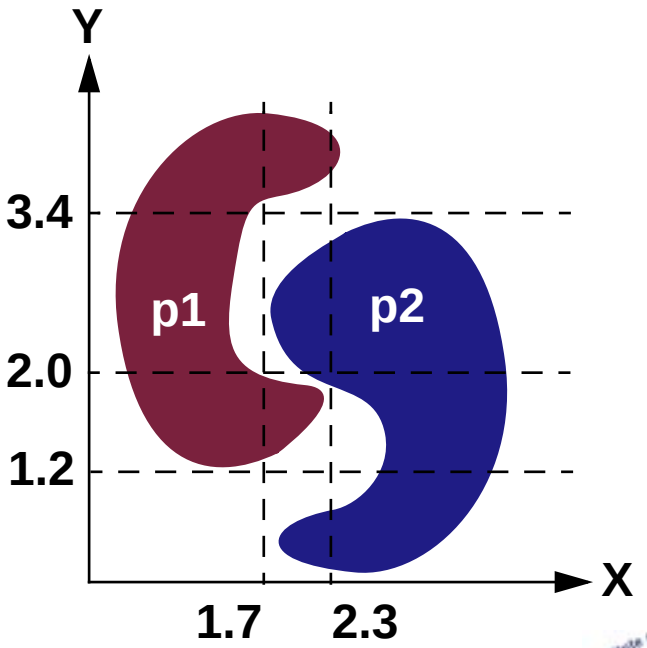


- Statistical neural networks fail to generalize on large data sets

Decision Tree Classifiers

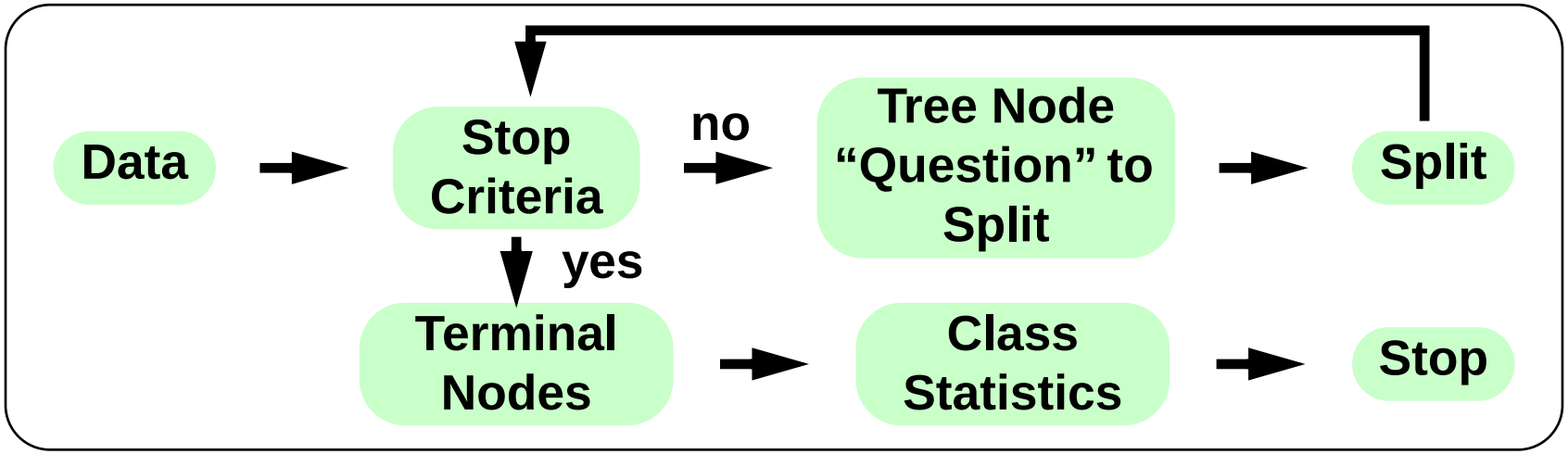
- Combine heuristics and statistics
- Ideally suited for nonlinear classification
- Classification based on data attributes





Decision Tree Overview

- Recursively partition data into groups



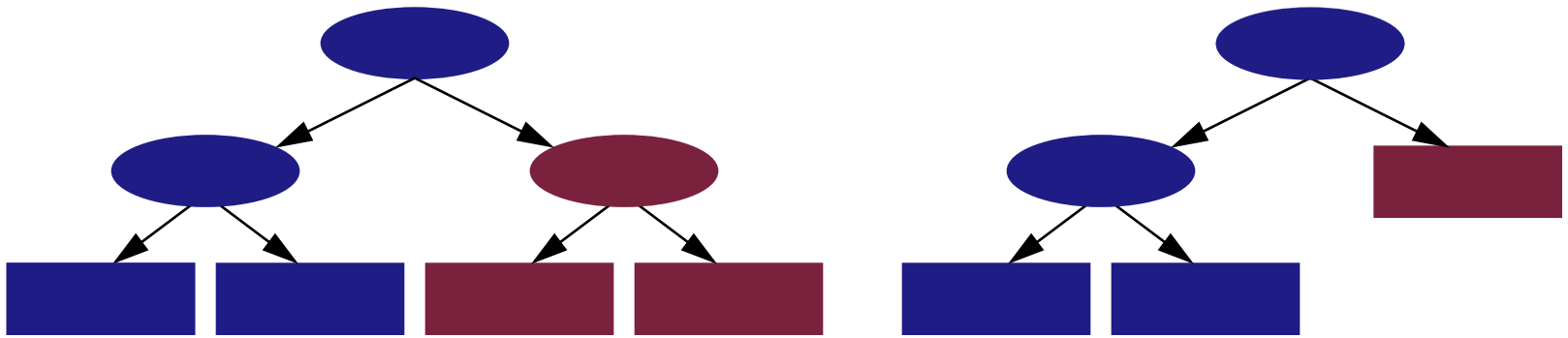
- Overfitting data — tree pruning
- Information theoretic basis for splitting, stopping, and pruning
- Capture complex relationships

DT Splitting Criteria

- Used to design "questions" at each node
- Bayesian splitting
  - — maximize *a posteriori* class likelihoods
- Information gain splitting
  - — maximize the information gain
- Information gain ratio splitting
  - — maximize information gain normalized by the total information in split
- Maximum entropy splitting

DT Pruning Criteria

- Collapse subtree into a terminal node if it results in lower predicted error
- Cost-complexity — trade off tree size and error rate
- Pessimistic pruning — statistical error estimates at node adjusted as per bias



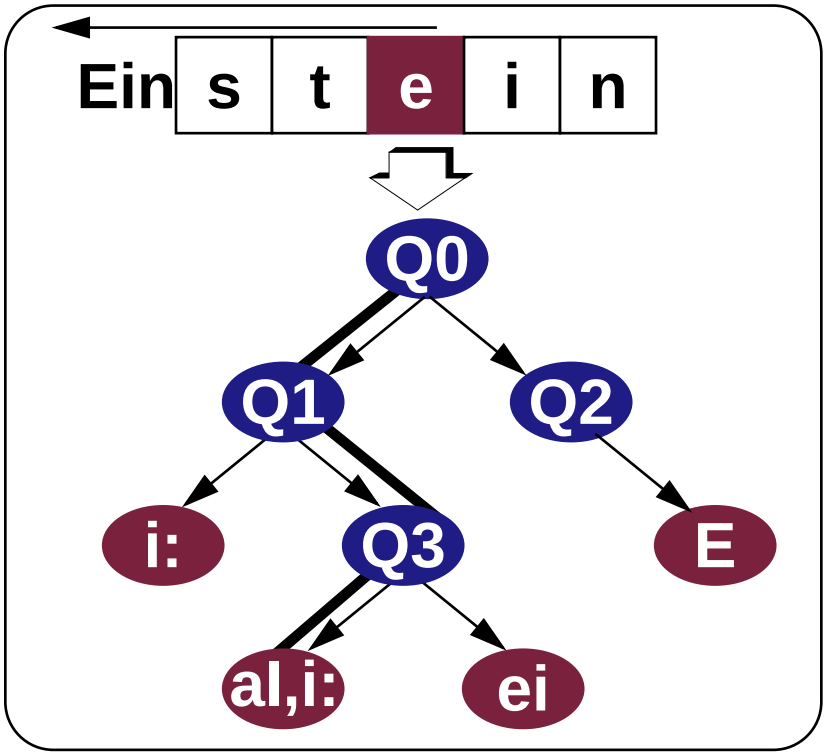
ISIP DT Toolkit

- Limitations of existing free software
  - — small number of classes
  - — number and values of attributes
  - — restrictions on class labels
  - — not amenable to text-based processing
- ISIP DT toolkit allows user-defined criteria for splitting, pruning, and stopping
- Data tagging allows selective attribute usage without reformatting

Surname Pronunciations

- Train DT on name-pronunciation pairs
- Sliding window of letter context

_	_	E	i	n	al
_	E	i	n	s	_
E	i	n	s	t	n
i	n	s	t	e	s
n	s	t	e	i	t
s	t	e	i	n	al
t	e	i	n	_	_
e	i	n	_	_	n



Pronunciation Dictionary

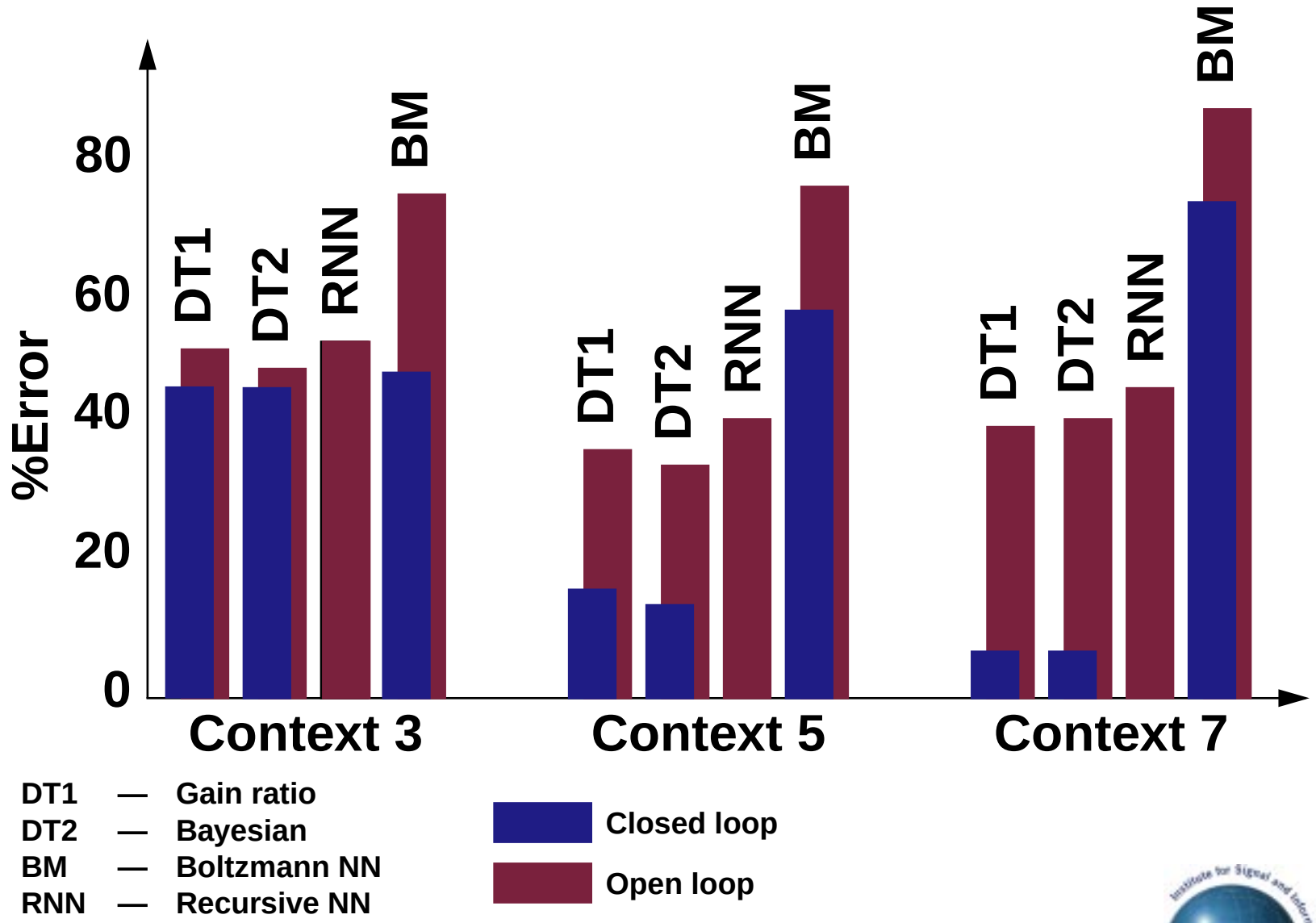
- 18494 surnames, 25648 pronunciations
- Worldbet phonetic convention to handle multilingual data
- Hand-transcribed pronunciations
- Automatic letter-to-phoneme alignment
- Public domain resource

Einstein	al _ n s t al _ n
Einstein	al _ n s t i: _ n
Fermi	f _ er m i:

Experiments

- Training set of 15000 names (3494 names held out for testing)
- Three such partitions for cross-validation of results
- Context lengths of 3, 5 and 7
- Binary univariate tree, single (1-best) pronunciation output
- Bayesian and gain ratio splitting
- Pessimistic pruning

Results



Institute for Signal and Information Processing

Analysis

- DTs try to model letter context-to-sound relationship
- As context size increases
  - — closed loop performance improves
  - — open loop performance decreases due to overfitting
- Extremely confusible classification space
- Performance of both splitting criteria is comparable
- Pruning only helps marginally

General English

- Used best surname-trained DT
- 8000 word subset from Switchboard lexicon as test data
- Word error rate 83%!!!
- Pruning makes no difference
- Letter-to-sound mapping for names is radically different from general English
- Issue — can a single tree be trained for both types of data?

Conclusions

- Decision tree systems perform better — 38% error
- Generation of multiple pronunciations will further improve error rate
- Highly nonlinear letter-to-phone maps — need more data for effective training
- Future plans — extend DT application to general English
- Public domain resources — DT toolkit, pronunciation dictionary

References

1. 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.

2. W. Buntine and R. Caruana. "Introduction to IND Version 2.1 and Recursive Partitioning," software manual, NASA Ames Research Center, Moffet Field, CA, 1992.

3. J. L. Hieronymus. "ASCII Phonetic Symbols for the World's Languages: Worldbet," *Technical Memo,* AT&T Bell Laboratories, 1994.

4. L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*, Wadsworth International Group, 1984.

5. W. Buntine, *A Theory of Learning Classification Rules*, Ph.D. thesis, University of Technology, Sydney, 1991.

6. J. R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, 1993.