KNOWLEDGE-GUIDED LEARNING OF STRUCTURAL DESCRIPTIONS

Mark S. Fox and !). Raj Reddy
Computer Science Department*

University
Pittsburgh, Pa. 15213

INTRODUCTION

We demonstrate how the use of domain dependent knowledge can reduce the combinatorics of learning structural descri.pt ions, using as *an* example the creation of alternative pronunciations from examples of spoken words. Briefly, certain learning problems (Winston, 19 70; Fox & Hayes-Roth, 19 76) can be solved by presenting to a learning program exemplars (training data) representative of a class. The program constructs a characteristic representation (Cl<) of the class that best fits the training data. Learning can be viewed as search in the space of representations. Applied to complex domains the search is highly combinatorial due to the: 1) Number of alternative CRs. 2) Size of training set. 3) Size of the exemplars.

.PROBLEM

An important aspect of speech understanding (Reddy, 19 76) research is the creation of wora pronunciation dictionaries. The dictionary entry for a word must contain the salient features and variations of how a word may be pronounced. The performance of a speech system depends directly upon the validity of this dictionary. Presently, a trained person constructs a dictionary by examining speech data, laboriously testing and modifying a dictionary many times. The goal of this work is to automate this process.

The training data is a set of segmented spoken words. Each exemplar is a varying length sequence of segments. A segment is a time partition of the speech signal that is labeled by an array of phoneme-weight pairs. The weight specifies now closely the segment matches the corresponding phoneme. The phoneme array for our application usually contains 100 entries and the average exemplar segment length is seven. Thus each exemplar depicts approximately 100 pronunciations (i.e., paths through the phoneme-segment matrix). The segment label's weight assignment is errorful. A segment is correctly classified (i.e.. correct label is best rated) only 42% of the time (Goldberg, 1976). Another problem is the segmentation process. An exemplar that contains possibly 10 segments may only represent 4 contiguous phonemes. Thus' all possible combinations (partitions) of contiguous segments must be constructed, and the labels that best describe the partitions must be determined.

The target CR is a network whose nodes represent phonemes and whose arcs represent allowable transitions between nodes. This network must be small due to space limitations.

The learning problem can be viewed as three separate phases: 1) Learning of a minimal state representation for each exemplar of a word class. 2) Combining the exemplars of a word class into a minimal .state network. 3) Modify the networks as a result of near miss analysis between word classes.

KNOWLEDGE-GUIDED LEARNING

Typical attempts at including knowledge in the learning process are generally ad hoc. We believe another important method of adding knowledge is to recognize where in the learning problem knowledge is critical in reducing the search space and emphasizing salient aspects of the exemplars. Then examine the domain for this critical knowledge. The importance of knowledge is primarily measured in terms of the reduction in search space. We believe that there exists a set of rules that can aid in the recognition and determination of this critical knowledge. The following Is a partial list of domain independent rules we have used to recognize relevant domain knowledge.

1) ".Bounds__on__RepresenLaLiQ": Information can narrow the minimum and maximum size bounds of the learned representation. For example: The number of syllabies (local maximum amplitude) In the example wave form provide a minimum size on the single exemplar learning.

2) Ff^t-nrP Focusing: Certain features ol an exemplar convey more Information than others and should appear in the final learned representation. For example: Acoustic characteristics of stressed syllables exhibit less ambiguity therefore are more reliable.

3) niffHrent-e " Focus1ng: It is important to emphasize features that differentiate between two similar classes (near miss analysis (Winston, 1970)). For example: Compare CRs and speech signal to derive differences.

A) .KcsnU CunfirmaLiun: There may exist a model of expectation of how the resulting CR should look. For example: 1) Each word can be described as a sequence of nasal, vowel, consonant, noise and silence type sounds.

AM EXAMPLE

Figure 1 depicts an exemplar of the word MENTION. Figure *1* depicts the best partition for two exemplars (single exemplar learning). The partition size was bounded below by the number of syllables and above by the abstract expectation model. The second exemplar was confirmed by the expectation. Figure 3 contains the resulting network constructed from the exemplars. Merging and separation of nodes was guided by both abstract models and signa1 1eve1 information.

DISCUSSION

Learning .in complex domains entails a large combinatorial search. The research described here attempts to solve two problems: Generally, where to look in a domain for knowledge that will help reduce the search space. Specifically, how to automatically generate word dictionaries for speech understanding systems both efficiently and accurately. At the time of writing the approach taken here appears to be fruitful. One idea that we see occurring repeatedly in this and other work (e.g.. vision) is the importance of models at various levels of abstraction in both the pruning and verification ot learned representations.

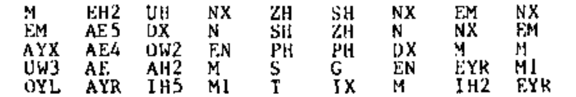
REFERENCES

Fox M.S. and F. Hayes-Roth, 19 76, "Approximation Techniques for the Learning of sequential patterns". Third Int. Joint Conf. on Pattern Recognition. Coronado Calif., Nov. 1976.

Goldberg, H.G., 19 76, "Segmentation and Labeling of Speech: A Comparative Performance Evaluation", (Ph.D. Thesis, Carnegie-Mellon Univ.), Tech. Report, Computer Science Dept., Carnegie-Mellon University, Pittsburgh PA.

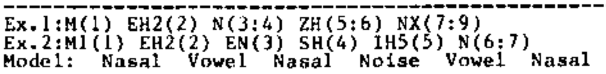
Reddy, D.R., 19 76, "Speech Recognition by Machine: a review". Proc. of the IEEE, (April 1976).

Winston P., 19 70, "Learning Structural Descriptions from examples", Tech. Report AI TR-231, MIT AI Lab., Cambridge, Mass..



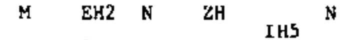
```
M    EH2  UH   NX   ZH   SH   NX   EM   NX
EM   AE5  DX   N    SH   ZH   N    NX   EM
AYX  AE4  OW2  EN   PH   PH   DX   M    M
UW3  AE   AH2  M    S    G    EN   EYR  M1
OYL  AYR  IH5  M1   T    IX   M    IH2  EYR
```

FIGURE 1: Example of the word MENTION.



```
Ex.1:M(1) EH2(2) N(3:4) ZH(5:6) NX(7:9)
Ex.2:M1(1) EH2(2) EN(3) SH(4) IH5(5) N(6:7)
Model:  Nasal  Vowel  Nasal  Noise  Vowel  Nasal
```

FIGURE 2: Highest rated partitioning for single exemplar learning. *(n:m) describes a partition where * denotes the best label that describes the partition, and (n:m) denotes the partition contains segments n through m. (Ex.1 is from Fig 1)



```
M    EH2  N    ZH        N
                    IH5
```

FIGURE 3: CR representing exemplars 1 and 2.

[1] This work was supported in part by the Defense Advanced Research Projects Agency under contract no. F44620-73-C-0074 and monitored by the Air Force Office of Scientific Research. In addition, the first author was partially supported by a National Research Council of Canada Postgraduate Scholarship.