Feature-Table-Based Automatic Question Generation for Tree-Based State Tying: A Practical Implementation

Supphanat Kanokphara and Julie Carson-Berndsen

Department of Computer Science
University College Dublin,
Ireland
{supphanat.kanokphara, julie.berndsen}@ucd.ie

Abstract. This paper presents a system for automatically generating linguistic questions based on a feature table. Such questions are an essential input for tree-based state tying, a technique which is widely used in speech recognition. In general, in order to utilize this technique, linguistic (or more accurately phonetic) questions have to be carefully defined. This may be extremely time consuming and require a considerable amount of resources. The system proposed in this paper provides a more elegant and efficient way to generate a set of questions from a simple feature table of the type employed in phonetic studies.

1 Introduction

Tree-based state tying technique is widely used to cluster HMM states into classes and tie all states in the same class in order to reduce the data sparseness problem [1]. The requirement for this technique is only a set of phonetic questions. While this strategy is good, poorly-defined phonetic questions may lead to lower accuracy in the resulting system. In order to use this approach to its full advantage, the phonetic questions must be defined by an expert who is familiar with the units and has a strong linguistic background. This may slow down the implementation of speech recognition systems since manual definition of phonetic questions is a time consuming task and, unless the data is thoroughly cross-checked, may be inconsistent and contain errors which may lead to degradation in the system.

Many researchers aware of this problem and have investigated alternative ways to generate questions automatically without any human intervention [2], [3]. The basic idea is to determine phone classes according to the database in a data-driven manner. However, the disadvantage of a data-driven approach is they might generate poor quality questions if the corpus is not of an appropriate quality.

To deal with the shortcomings of the manual and data-driven approaches simultaneously, we suggest a separation of the question generation procedure into 2 different steps, namely *feature tagging* and *feature co-occurrence tagging*. Feature tagging is the process of examining the relationship between a unit (in this case, phone) and its corresponding features. This process has two possible outputs: classes of units defined

according to their features or units tagged with their respective features. Feature co-occurrence tagging is the process of examining how features overlap (or co-occur) and defining classes of units which model the co-occurring features. For example, in English, a lip rounding feature can co-occur with a vocalic manner feature but not with a stop manner feature. The feature co-occurrence tagging step is carried out automatically given the tagged feature set. By doing this, the requirement for linguistic experts for phonetic questions is certainly reduced; in some cases the linguistic expert may not even be necessary because feature tagging is quite common in linguistics and thus tagged feature sets are already available for many languages in the form of feature tables [4], [5]. This novel approach addresses the shortcomings mentioned above since feature tagging is based entirely on linguistic knowledge and hence robust to bad quality corpora.

2 Feature-Table-Based Automatic Question Generation

Due to space limitations, the algorithm will not be fully explained here but interested reader can find it from [6]. In [6], we generate all possible feature co-occurrence classes and prune linguistically ill-formed classes later. This is considered to be computational inefficiency because many linguistically ill-formed classes have to be constructed. In this paper, we introduce another tree-based clustering to generate feature co-occurrence classes. This allows us to prune out some classes while they are constructed. Moreover, when a node is pruned, its entire child nodes are also pruned thus reducing system complexity.

The tree is constructed in a left-to-right, top-down fashion. All of the nodes on a particular level are expanded before moving down to the next level. For the purposes of this paper, we assume that each node of a decision tree is a feature co-occurrence class and every leaf node is a linguistically well-formed class. The depth of a tree is equal to the number of tiers (i.e. a particular level in the tree represents a specific tier) and the number of branches for each node is equal to the number of features on the next tier. The tree expansion continues until tier *N* is reached and nodes which remain at tier *N* are assumed to be linguistically well-formed classes.

It is important to note that this tree is not the same as tree-based state clustering. Tree-based state clustering forms a phone set according to the probability score and a question at each node is chosen in maximum likelihood sense. Our tree clusters a phone set orderly according to a feature table. In tree-based state tying, a question (phone class) for a child node is a subset of its parent node question, i.e. liquid → l, etc. In our system, a phone class of a node does not have to be a subset of its parent node. This allows our tree to construct feature co-occurrence phone class automatically.

Actually, the phone recognition results from the algorithm in this paper and [6] are the same. The difference is just time for building phonetic questions. Phonetic questions can be constructed much faster than the ones in [6]. Therefore, we expanded our feature table to include gender tier. With this gender tier, we can expand our acoustic

model to be gender-dependent. This increases phone recognition accuracy from 71.14% to 72.49%.

3 Conclusion

This paper has proposed a novel way to generate a set of questions for tree-based state tying. This strategy requires only a simple feature table which is likely to available in many languages since these are commonly used for phonetic and phonological studies. This system is very convenient where a speech recognition system has to be developed.

Since an extra tree clustering is introduced in this paper, phonetic questions can be generated faster and more efficient. This allows us to include gender information in our feature table and model gender-dependent acoustic model. This gender-dependent model and our phonetic questions show better phone recognition accuracy.

Acknowledgements

This material is based upon works supported by the Science Foundation Ireland for the support under Grant No. 02/IN1/I100. The opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of Science Foundation Ireland.

We also would like to thank Dr. Lorraine McGinty for AI aspect discussion and Mr. Moritz Neugebauer for reviewing our paper.

References

1. Odell, J.J.: The Use of Context in Large Vocabulary Speech Recognition. Ph.D. Thesis. Cambridge University, Cambridge (1995)
2. Beulen K., Ney H.: Automatic Question Generation for Decision Tree Based State Tying. in Proc. Int. Conf. Acoust., Speech, Signal Processing, Vol. 2 (1988) 805-809
3. Singh, R., Raj, B., Stern, R. M.: Automatic Clustering and Generation of Contextual Questions for Tied States in Hidden Markov Models. in Proc. Int. Conf. on Spoken Language Processing. Vol. 1 (1999) 117-120
4. Geumann, A.: Towards a New Level of Annotation Detail of Multilingual Speech Corpora. in Proc. Int. Conf. on Spoken Language Processing. (2004)
5. Luksaneeyanawin, S.: Speech Computing and Speech Technology in Thailand. in Proc. The Symposium on Natural Language Processing. (1993) 276-321
6. Kanokphara, S., Geumann, A.,Carson-Berndsen, J.: Accessing Language Specific Linguistic Information for Triphone Model Generation: Feature Tables in a Speech Recognition System. Submitted to 2nd Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics. (2005).