論文 / 著書情報
Article / Book Information

論題(和文)	
Title(English)	Markov Model Reordering of Sentence Hypotheses
著者(和文)	古井 貞熙
Authors(English)	Michael Barlow, Tatsuo Matsuoka, Sadaoki Furui
出典(和文)	, , , pp. 175-176
Citation(English)	1995年日本音響学会春季講演論文集, , , pp. 175-176
発行日 / Pub. date	1995, 3

MARKOV MODEL REORDERING OF SENTENCE HYPOTHESES *

Michael Barlow, Tatsuo Matsuoka & Sadaoki Furui †
(NTT Human Interface Laboratories)

1 Introduction

The ATIS (Air Travel Information System) task represents many of the most difficult problems faced today in speech recognition research. A speech understanding system that deals with spontaneous, continuous speech from a number of speakers is required. An important aspect of any such speech understanding system is the grammar modeling.

A common approach to speech understanding tasks is to break the system into two major components. The front-end being a speech recogniser that incorporates limited grammar modeling (e.g., bigrams or trigrams). The back-end being the speech understanding component that translates the top hypothesis generated by the speech recogniser into the required response (in the ATIS task generating an SQL query of the database).

This paper describes the design of an intermediary component in the speech understanding system between the speech recogniser and the understanding component. The intermediary stage serves two purposes. Firstly it incorporates more sophisticated grammar modeling than that found in the speech recogniser. As such it can be used to rescore and reorder the hypotheses generated by the speech recogniser. Secondly the grammar used can also provide a partial semantic breakdown of the sentence which can be employed by the understanding component.

In particular this paper will describe a Markov Model based grammar model and its employment to rescore/reorder the hypotheses generated by a speech recogniser.

2 Method

An N-best tree-trellis[1] based speech recogniser was run on a set of 455 sentences from the ATIS-2 domain (see data description). The top 25 hypotheses generated by the speech recogniser for each sentence were then examined. If the hypothesis was considered semantically equivalent to the original query (from an SQL query generation definition) then the hypothesis was marked as acceptable. For example the sentences: "Find flights from Pittsburgh to...", "Find flights Pittsburgh to..." and "Find a flight for Pittsburgh to..." were all considered equivalent.

A second set of 912 transcribed (text) sentences were used in the training of the Markov models and Hidden Markov models for the ATIS grammar. These sentences were hand labeled, in a method similar to Pieraccini et. al. [2], each word in each sentence being assigned to one of 53 classes based on the semantic function of the word within the sentence. These classes fall into three major categories:- Restrictions, e.g., airline ("...on U.S. Air...") or departure time ("...leaving before 4 p.m...."), Attributes, e.g., fare ("What is the cost...") or aircraft ("Which aircraft fly between...") and a third class containing such things as invocation ("Give me..."), action ("Reserve a limousine..."), and joiner ("...and...").

These classes were then equated with Markov Model (MM) states, yielding a 53 state ergodic MM to describe the ATIS grammar. Due to the limited amount of training data MMs with a smaller number of states were also considered by combining multiple classes into a single state. By this method 53, 46, 12 and 3 state MMs were considered.

The same data (ignoring the labeling) was also used to train (Baum Welch re-estimation) the hidden Markov models. Due to the limited training data only 3 and 5 state HMMs were considered.

A simplified form of bigram output probabilities was considered for the HMM/MMs due to the limited training data. The speech recogniser employed a lexicon of 1279 words. Each word was assigned a category corresponding to its *Parts of Speech* with the addition of several special classes corresponding to the ATIS task (e.g., location and time), yielding a total of 18 categories. The output probabilities for a state were then of the form: P(<word> | <category of previous word>).

Consider:
$(w_1, w_2, w_3, \ldots, w_N)$ word sequence
$(p_1, p_2, p_3, \ldots, p_N)$ corresponding word category sequence
$(c_1, c_2, c_3, \ldots, c_N)$ corresponding semantic categories
then
$P(w_i|p_{i-1}, c_i)$ emission probability
$P(c_i|c_{i-1})$ transition probability

The labeled training data was then used to exactly calculate the above probabilities for the different Markov Models (different number of states) examined.

As well as the pseudo bigram output probabilities ($P(w_i|c_i, p_{i-1})$), unigram output probabilities ($P(w_i|c_i)$) and a further simplified bigram output probability ($P(p_i|c_i, p_{i-1})$) were also considered. In addition a form of global smoothing (state independent smoothing) was applied to both the output and initial state probabilities due to data limitations:-

$$\alpha P(w_i|c_i, p_{i-1}) + (1-\alpha)P(w_i|p_{i-1})$$

$$\beta\pi_i + \frac{(1-\beta)}{NS}$$

where α is the output smoothing factor, β is the initial state smoothing factor, and NS is the number of states used in the model.

Finally, in order to test the models the trained Markov Model was used to score the top 25 hypotheses generated by the speech recogniser for each of the

*マルコフモデルによる文仮説の再編成　†バーロウ・マイケル　松岡達雄　古井貞熙　(NTT ヒューマンインタフェース研究所)

455 testing sentences. The likelihood scores of the recogniser and grammar components were combined and the hypothesis selected that had the highest likelihood score. This selection was then checked against the list of acceptable (semantically equivalent to original utterance) hypotheses for that sentence.

$$\arg\max_i P_R(h_i)P_G^N(h_i)^w$$

where h_i is the i'th hypothesis, $P_R()$ is the recogniser probability, $P_G^N()$ is the normalised grammar probability and w is the weighting factor.

3 Data

Two separate sets of data were used for training and testing of the system. For training the 912 sentences composing the ATIS-0 test and training set were used. For testing 455 sentences from that ATIS-2 database were used. These comprised 194 sentences from the ATT subset and a further 261 sentences from the CMU and SRI subsets.

4 Results

Table 1 shows the best results obtained by combining the Markov Model ATIS grammar with the speech recogniser scores to reorder the sentence hypotheses. These results were obtained using a 53 state Markov model with output probabilities of the form $P(w_i|c_i, p_{i-1})$.

Data	*Recog.*	*+M.M.*	*Improved*
ATT(194)	41.8%	46.4%	4.6%
CMU+SRI(261)	69.7%	72.4%	2.7%
All(455)	57.8%	61.3%	3.5%

Table 1: Results of MM based reordering of Speech Recogniser Sentence Hypotheses with absolute improvement.

From the table it can be seen that a small but significant improvement (3.5%) in selection of an acceptable sentence occurs if the MM based grammar model scores are combined with the speech recogniser scores. The improvement is more marked for the ATT collected speakers (4.6%), who have in general a very poor correct recognition rate (below 50%). On the other hand the improvement for the CMU and SRI speakers is less marked (2.7%).

It is worth noting that in many cases the speech recogniser produced no acceptable hypothesis amongst its top 25 candidates. In such a circumstance the addition of the hypothesis processing grammar model *cannot* improve performance. In particular 38.7%(75) of the ATT sentences, 9.6%(25) of the CMU/SRI sentences and 22.0%(100) of all sentences could not be correctly hypothesised. If these 'no-acceptable-hypothesis' sentences are eliminated in measuring the performance improvement achieved by the incorporation of the MM scores, then a 7.5% improvement was obtained for the ATT data, a 3.0% improvement for the CMU/SRI data, and an overall 4.2% improvement.

Results for the HMM based grammar model showed no improvement over the baseline speech recogniser performance (rescoring with HMM did not improve performance). This appears attributable to the lack of training data.

Results for the different output probabilities considered consistently showed that output probabilities of the form $P(w_i|c_i, p_{i-1})$ were the best, followed by unigrams and then the extremely simplified bigram ($P(p_i|c_i, p_{i-1})$). Recognition performance was also affected by the number of states with the larger number of states yielding better performance.

The choice of the smoothing parameters α and β also affected performance. The optimal parameters were found to be $\alpha = 0.95 - 0.98$ and $\beta = 0.1 - 0.5$. This indicates that the output probabilities need little if any smoothing while the initial state probabilities need considerable smoothing.

5 Discussion

The above results have extended the work of Pieraccini et. al.[2]. Pieraccini et. al. considered a limited domain of the ATIS task (class A sentences) and only applied their grammar models to the segmentation of new data. As an extension this work considers all sentence classes and applies the grammar model as a post-processing adjunct to the speech recogniser for improving selection of an appropriate sentence hypothesis.

The results for the MM/HMM grammar model show considerable scope for further improvement, as the current system yields only a small improvement in overall recognition rate.

A possible cause of this relatively low improvement in performance may be inadequate training data. Indeed this is highly likely as the training data consists of roughly 10000 words (tokens) and in many cases more MM/HMM parameters than that are required to be calculated (though, due to the specialised nature of the MM states it is reasonable and meaningful for most of the output matrix to be null). On this note also it would be reasonable to expect that full bigram output probabilities within a state would yield higher performance. However this corresponds to a 50-fold increase in number of output parameters and hence significantly more training data is required.

References

[1] Chou, W., Matsuoka, T., Juang, B.-H., Lee, C.-H., "An Algorithm for High Resolution and Efficient Multiple String Hypothesization for Continuous Speech Recognition using Inter-Word Models", Proc. ICCASP-94, vol 2, 153–156, 1994.

[2] Pieraccini, R., Levin, E., Lee, C.-H., "Stochastic Representation of Conceptual Structure in the ATIS Task", Proc. DARPA Speech & Natural Language Workshop, Monterey, CA, 121–124,1991.