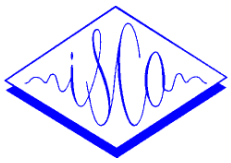

3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

DATA-BASED CONTROL OF THE SEARCH SPACE GENERATED BY MULTIPLE KNOWLEDGE BASES FOR SPEECH RECOGNITION

Kevin Johnson, Roberto Garigliano and Russell J. Collingham

Laboratory for Natural Language Engineering
Department of Computer Science, University of Durham, England, DH1 3LE

Abstract

Automatic speech recognition (ASR) systems are made up of a number of different knowledge sources (KSs) which combine to solve the overall problem of speech recognition. An investigation into the benefits of each KS and the benefits of combining the KSs has been undertaken. Test sentences, taken from real data, have been processed by different versions of a speech recognition sub-system. The system performance was measured accurately by identifying the exact location of the required hypothesis (RH) in the hypothesis list (HL), at all cycles in the processing. This paper describes the analysis carried out and presents the results of this analysis and shows that each KS does play a role, though their importance varies considerably. The overall aim of this work is to aid in the development of a speech recognition system being produced at the University of Durham.

1 Introduction

The fast, incremental nature of the development of ASR systems is resulting in systems comprising of multiple KSs with each KS performing a specific task within the system. The question that needs to be asked is which KS actually benefits the system and how the different KSs cooperate to achieve the overall goal of the system. Using the sub-system of an ASR system, currently being developed at the University of Durham, an investigation of each KS used by the system has been undertaken. Each KS was investigated to identify the advantages and disadvantages of using it and the effect on the system's overall performance when the KS is used. The original sub-system was modified so that several versions of the system could be easily created. Each version processed ten test sentences and measurements were taken on the search space generated during each run and performance of each system. This allows the best system to be identified and the bottom line performance of each individual KS, both alone and in co-operation with other KSs, to be identified. The aim of the analysis was to decrease the search space of the ASR system, thus increasing the time performance, without diminishing the accuracy of the system.

2 Current System

The system, [1] at present, is only part of a final product. The front end is a simulated phoneme recogniser with a corruption rate of up to 25%. The corruptions are split between insertions, substitutions and deletions based on the details given in [2]. The ASR system (AURAID - A University speech Recognition AID) uses the first level of a two level dynamic programming algorithm to build a word lattice and a pyramid beam search with a skip and share algorithm to determine sentence hypotheses occurring in the lattice. Word frequency information and anti-grammar (AG) rules are then used to select the most appropriate hypothesis. The skip and share algorithm deals with the insertion and deletion of phonemes while the word frequency processing uses information from the Oxford Advanced Learners Dictionary [3] to categorize words as common, normal or rare. The AG rules [1] are a set of rules which indicate what cannot be said or more to the point how things are not normally said during spontaneous natural speech. It is a set of rules which decreases the likelihood of hypotheses being selected and therefore removes ill-formed hypotheses from a list of potential hypotheses.

3 Data Preparation

A number of lectures, performed at the University of Durham, were recorded and transcribed to include all of the disfluencies of speech, including the "hums" and "ah" as well as pauses. The data used in the analysis presented in this paper was taken from a single lecture on "Software Engineering" given as the first introductory lecture for second year students. The lecture contained 4903 words and 382 sentences, or part sentences, with an average of 12.84 words per-sentence. From this lecture 10 representative sentences were selected. They were representative in sentence length and speech disfluencies, which are of great interest to the authors [4] therefore 5 of the sentences contained repairs. Two dictionaries were used in the analysis. The first contained 528 words of which 354 were from the LOB corpus and 146 were category words which are important to the field of lectures. The second contained 1985 words and was used to test the systems performance on a more realistically sized vocabulary.

4 Analysis

The aim of the analysis was to decrease the search space thus increasing the time performance, without diminishing the accuracy of the system. Ten representative sentences were taken from a recorded lecture. Eight versions of the system processed each sentence and the results were compared.

Switches were built into the original system to allow different version to be easily created. Eight different switches were used:

1. Unrestricted beam width
2. Highly restricted beam width (10% & 5%)
3. Medium beam width (20% & 10%)
4. Skip and share algorithm
5. Word frequency information
6. Anti-grammar rules
7. 500 word dictionary
8. 2000 word dictionary

The combinations of the switches which made up the eight system can be seen in Table 1.

The data collected on each run included: system time; elapsed times; cycle times (a cycle is the inclusion of a word in the hypotheses); number of hypotheses expanded on each cycle; number of hypotheses generated on each cycle; position of the RH in the HL; hypothesis score of the RH; the cut off score (score beyond which no hypothesis was expanded); error rate; word accuracy; words correct; the percentage position of the RH from the top of the HL (i.e. 10% from the top of the list) and the percentage score of the RH from the top (i.e. the percentage difference between the score of the top hypothesis and the score of the RH).

System	Switches							
	1	2	3	4	5	6	7	8
1	x						x	
2		x					x	
3			x				x	
4			x	x			x	
5			x		x		x	
6			x	x	x		x	
7			x	x	x	x	x	
8			x	x	x	x		x

Table 1: Systems analysed

It must be noted that in identifying the position of the RH (the actual input) within the HL the type (VERB, etc.) of the word was used as well as the word itself along with the exact phoneme location. This makes the details very accurate and the figures seem lower than those systems whose performance is measured on word identification alone.

The results were compared to see if the included KS had any effect and whether the effect, on combination with other KSs, was beneficial to the system as a whole.

A subsequent manual analysis on semantics and repair was undertaken to investigate the performance and accuracy of possible expansions. This manual analysis consisted of looking through the HLs for three of the test sentences and identifying the outcome of each of the sources of knowledge by manually scoring those hypothesis that satisfied certain criteria. An improved position of the RH, in the HL, would show that the inclusion of some sort of KS on the topic investigated would be beneficial.

5 Results

The results can be split into five different sections. The first looks at the effect of changing the beam width(s of the pyramid beam search . The second section looks at the inclusion of a skip and share algorithm and word frequency information for sentence selection. The third section looks at the inclusion of anti-grammar (AG) while the fourth section looks at the effect of increasing the dictionary from 528 words to 1985 words. And finally the fifth section compares the overall performance of each systems.

5.1 Beam width

A pyramid beam search, where the width for early cycles is wider than the rest, is used to ensure that the RH is not lost early in the processing, before a more detailed analysis could be done. The first thing to investigate was the actual width(s) used in the beam search. The beam widths investigated include an unrestricted; narrow and medium beam width.

The system, with an unrestricted beam width, resulted in an astronomical search space which took many hours to process. After three hours of processing the first sentence, the system was stopped, still running on cycle four (increasing the hypotheses to four words in length) and had created 82,066 hypotheses. A narrow beam width resulted in a more acceptable performance but still resulted in slow processing times. This was mainly due to the RH being outside the initial beam width and therefore not expanded leaving incorrect hypotheses which were expanded and created further incorrect hypotheses of similar scores. A medium beam width resulted in the RH being expanded and allowed a few hypotheses to have increased scores. Thus the system expanded these few hypotheses and the system's time performance increased. It was found that a medium beam width was satisfactory in that it covered the RH and resulted in respectable performances, therefore a wider beam width was not investigated and a medium beam width was used in the rest of the systems.

5.2 Skip & Share and Word frequency

The introduction of skip and share processing made very

little difference to the overall performance of the system. It did not increase the systems performance though it did show promise in overcoming one of the problems of repair by bridging a part word, but the resulting string had such a low score that it was never expanded.

Word frequency information gave a definite increase in system performance for both system time and position of the RH. Though not producing a completely satisfactory result it did go some way to moving the RH to the top of the hypotheses list.

A combination of skip and share processing and word frequency information showed a slight increase in performance over the word frequency information alone but this was mainly a time increase rather than a performance increase. This system, with a combination of skip and share processing and word frequency information, was taken as the basis for the rest of the analysis.

5.3 Grammar

The system, using AG rules, worked much better than the other systems (i.e. those without anti-grammar) as the RH was generally higher in the hypotheses list, though it was not necessarily top of the list. This is not a major problem to this work as the accuracy of the measurements are such that a higher position in the list is more desirable as it shows that the extra KS is, in fact, being beneficial. One problem identified with the AG is that it did not cope well with repairs which is not surprising as it was not designed to cope with repairs.

5.4 Increased dictionary

The performance of the system using a more realistic vocabulary of 1985 words was acceptable. The systems performance did not decrease as would be expected but increased for all sentences. The changes is system time fluctuated across the test sentences (some increased and some decreased) but the position of the RH in the list of hypotheses generally increased.

5.5 System performance

Generally the system showed an increased performance both in system time and position of the RH when KSs were combined. This can be seen in Figure 1 which shows the percentage the position of the RH, for four of the systems analysed. Though the actual score difference for the RH does not chance much the important thing to note is in figure 1 where system 7 shows a definite position increase. It must also be noted that system 7 expanded the RH onto word seven which was not done by any other system. This is also true for system 6 which shows a better performance than system 3. This example, taken from the ten test sentences shows that combining KSs is helpful.

As well as this information word accuracy for each system was also calculated. This measure of accuracy was not deemed as important as the HL measurements as this research was interested in the progress of the RH when knowledge was added to the system. Table 2 shows

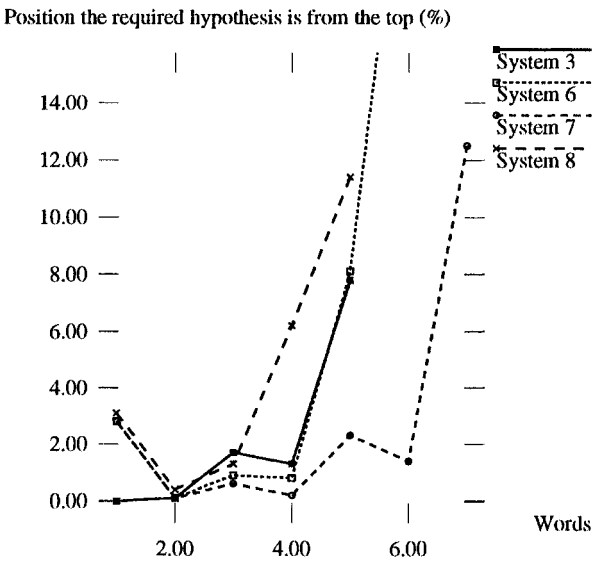

Figure 1: Percentage position for sentence 7

that the word accuracy, though generally high, fluctuated from system to system and does not give any real data on the performance of each KS. Therefore the above measure on hypothesis score and position are much better for the analysis presented here.

6 Problems

There were two main problems found with the system. The first is that repairs cause a major problem with the AG. AG has been produced to punish those hypotheses that have ill formed structures. It is well known that repairs [4] [5] [6] are a definite problem found in spontaneous speech and it is expected that repairs do not follow normal constructs of natural speech. As repairs are ill-formed structures and AG punishes ill-formed structures it is necessary to have something that deals specifically with repairs A follow-up analysis has been undertaken [7] to look at the true effect of repairs.

A further problem, and one of the main reasons why this research was undertaken, is the search space generated by the system. A limit on the number of hypotheses created at each cycle/expansion of the hypotheses along with the beam width restricts the search space but it can cause problems. When the limit is low the RH may not be expanded as higher hypotheses may create the limited number of new hypotheses while a high limit may increase the search space dramatically. The only way of overcoming this problem is to ensure that the RH is as near to the top of the HL as possible. This is where the inclusion of extra KSs is necessary.

7 Manual Analysis

In an attempt to predict the effect of extra KSs a manual analysis of semantic and repair processing has been undertaken. For semantics three of the test sentences have been taken and those cycles that contained

sent.	System						
	2	3	4	5	6	7	8
1	53.8	61.5	69.2	53.8	53.8	61.4	46.1
2	61.5	53.8	46.1	61.5	61.5	53.8	53.8
3	38.4	46.1	46.1	46.1	53.8	46.1	46.1
4	00.0	66.6	66.6	58.3	58.3	66.6	50.0
5	83.3	91.6	75.0	75.0	75.0	83.3	83.3
6	58.3	58.3	58.3	50.0	50.0	75.0	50.0
7	100	100	100	91.6	91.6	100	91.6
8	72.7	72.7	54.5	72.7	63.6	72.7	54.5
9	0	63.6	63.6	54.5	54.5	54.5	54.5
10	70.0	70.0	70.0	70.0	70.0	80.0	80.0

Table 2: Word Accuracy

three or more words have been examined. Each hypothesis was scored based on its meaning potential. The HL was then re-ordered and the position of the RH noted. The average increase of 32.79%, in the position of the RH, shows that semantic knowledge is potentially beneficial. An analysis of two of the test sentences has shown us that repair analysis would not increase the search space dramatically. Only 7.66% of hypotheses, that contained two or more words, contained possible repairs. If these were overcome the semantics processing would improve the position of the RH.

This manual analysis though not conclusive does show that the inclusion of extra KSs on semantics and repair would be beneficial to any ASR system.

8 Conclusion

This paper presents an analysis of a sub-section of an ASR system being developed at the University of Durham. Eight versions of the system, created by switching on or off different sources of knowledge, processed ten test sentences taken, as representative sentences, from a lecture presented at the university of Durham.

The measurements taken were very accurate, using the exact word pert of speech tags (VERB, etc.) to identify the location of the RH within the HL. The results of the eight systems were compared to see the effect of each KS. The results showed that a medium beam width (20% and 10%) was best and word frequency information and anti-grammar rules were also very beneficial but a skip and share algorithm was not. It also showed that using an increased vocabulary was not a detriment to the system. Problems were noted on processing repairs which need to be overcome but a further manual analysis showed that extra knowledge on repairs and semantics could overcome this problem.

This work gives us confidence in the system and shows that extra work on the skip and share algorithm is required and that it is possible to build on the current system knowing that there is a sound foundation. The production of a repair module and semantic module is the next task for the project.

9 Acknowledgements

Kevin Johnson was funded by a Science and Engineering Research Council CASE Studentship in collaboration with the Speech Research Unit of the Defence Research Agency, Malvern. Our thanks must also go to CSELT in Italy without whose help this paper would not exist. The views expressed in this report are those of the authors and not necessarily of the University of Durham, SERC, the DRA or CSELT.

References

[1] R. J. Collingham and R. Garigliano. Using anti-grammar and semantic categories for the recognition of spontaneous speech. In *Proceedings of Eurospeech, the 3rd European Conference on Speech Communication and Technology*. ESCA, September 1993. Berlin.

[2] S. R. Browning, R. K. Moore, K. M. Ponting, and M. J. Russell. A phonetically motivated analysis of the performance of the ARM continuous speech recognition system. In *Proceedings of the Institute of Acoustics Speech and Hearing Conference*, November 1990. Windermere.

[3] R. Mitton. *A Description of a Computer-Usable Dictionary File Based on the Oxford Advanced Learner's Dictionary of Current English*, June 1992.

[4] K. Johnson, R. J. Collingham, and R. Garigliano. Data-supported case for the extended coverage of repairs in the recognition of natural speech. In *Proceedings of the Institute of Acoustics Autumn Conference : Speech and Hearing*, November 1994. Windermere.

[5] J. Bear, J. Dowding, and E. Shriberg. Integrating multiple knowledge sources for detection and correction of repairs in human-computer dialog. In *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics*, pages 56–63, June 1992. Delaware, USA.

[6] J. Hirschberg and C. Nakatani. A speech-first model for repair identification in spoken language systems. In *Proceedings of the 3rd European Conference on Speech Communication and Technology (Eurospeech'93)*, volume 2, pages 1173–1176, September 1993. Berlin, Germany.

[7] K. Johnson, R. Garigliano, and R. J. Collingham. The effect of repair on speech recognition performance. In *Submitted to the 4th Conference on Applied Natural Language Processing*, October 1994. Stuttgart, Germany.