ON MULTI-LEVEL MACHINES FOR CONTINUOUS SPBBCH RECOGNITION

Joseph di Nartino

UNIVBRSITE DB NANCY I - CENTRE DE RECHERCHE EN INFORMATIQUE DE NANCY

ABSTRACT

In this paper we introduce the concepts of multi-level machines and of multi-level dynamic programung. These machines are well suited for the difficult continuous speech recognition problem because firstly, they permit the integration of several knowledge sources, secondly, they allow an optimal search based on dynamic programmng and thirdly, they can deal with semantic constraints. Furthermore these semantic constraints have the interesting property to be dynamic, i.e. they can be modified easily by the speech recognition system itself. The important consequence of this property is that the multi-level machines presented in this paper have a potential self-leaminy hability.

i - iroroacricN

In this paper we present the concepts of multi- level machines and multi-level dynamic programming. These muulti-level machines, on one hand, permit the integration of several knowledge sources such as phonology, syntax, semantic etc, and on the other-hand can take into acount local semantic constraints. As it will be shown these semantic constraints are "dynamic", in the sense that the speech recognition system can modify them easily.

In section 2 we begin by showing how a 1-level machine can be built from simple finite-state automatons. These automatons are called cells because the entire machine is built from these elementary machines. The description is done in such a way that the iterative process to generate a general n-level machine can be induced easily. The semantic links are also describes and we put in evidence the fact that firstly, they can be modified easily, and secondly that such modifications can be realized by the speech recognition system itself. This interesting property confer to the system a self-learning hability.

In section 3 we introduce the formalism necessary to explain how an optimal seauch can be realized in an n-level machine. From this mathematical discussion we show that the solution found is optimal for all the levels of the machine.This other intersting property confirms the power of the n-level machines.

II - DESCRIPTION OF THE MACHINE

The basic of the machine is a simple finite-state machine as illustrated by figure 1 (1).

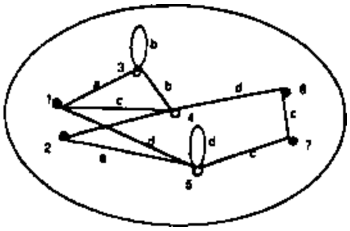

FIGURE 1. The basic cell of a multi-level machine : a simple finite state automaton.

The basic cell is characterized by a set of starting states ES, a set of ending states FS, and a set of terminal symbols T. For example, in the case of figure 1, $ES = (s_1, s_2)$, $FS = \{s_6, s_7)$ and $T = (a,b,c,d)$. An element of the starting state set and an element of the ending state set are particularized : they are called respectively the entry-state of the machine and the exit-state of the machine. These notions of entry-state and exit-state will be used later on when the notion of context of a machine is defined.

Now in order to connect these basic cells we use a finite-state formalism in which the symbols are the basic cells we just described. Figure 2 shows a machine built from the connection of basic cells.

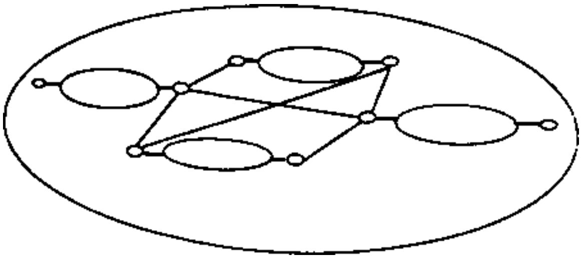

FIGURE 2. An example of a 1-level machine.

Let e_c and f_c designate respectively the entry-state and the exit-state of an elementary cell C.
Let n_c and m_c two nodes of the finite-state machine in which are embedded the elementary cells, defining a transition labelled by symbol C.

We define the context t^1 of the elementary cell C by :

$$t^1 = (n_c, e_c, C, f_c, m_c) \qquad (1)$$

Let $t'^1 = (n_{c'}, e_{c'}, C', f_{c'}, m_{c'})$ the context of another machine C'. Then we say that machine C is connected to machine C if the contexts of the two machines are tied by the relation of partial order defined by relation (2) :

$$t^1 \leqslant t'^1 \text{, iff } m_c = n_{c'} \qquad (2)$$

Relation (2) defines the connection of two machines. Bat, in fact this link is strengthened by a special link which ties the exit-state of one machine with the entry-state of the other one. we call such a link a semantic link because by specifying a single node of the target machine a lot of paths-semantioally incorrect- are suppressed. Furthermore as these semantic links can be modified dynamically in a straighfbrward manner by the speech recognition system, we claim that the multi-level machines have a self-learning hability. Figure 3 illustrates such a semantic link.

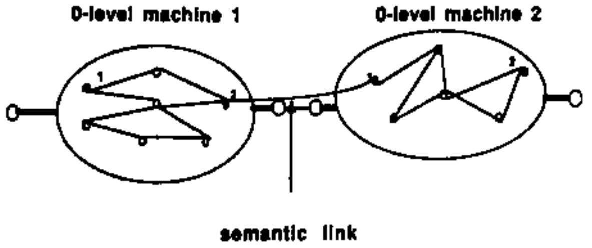

FIGURE 3. An illustration of a semantic link in a 1-level machine.

At this stage of the developpement we can give the following definitions :

- The basic cell is called a Olevel machine. Such a machine can be used for example to represent the acoustical, phonological variations of a lexical entity. Interesting studies on this subject can be found in 5 and 6.

- A machine built from basic cells is called a 1-level machine.

Figure 2 is an example of a 1-level sachine.

Now, it is clear that our procedure for building a 1-level machine from Olevel Mchines can be iterated. We get in this case a general n-level menine ; Such machines are interesting because of the following properties :

- 1 they can tackle the different degrees of abstraction of the language ; in other words the structure of the language can be dealt with by these machines. This property is a very desirable one, because the languages represented in not a structural way -i.e. for example Olevel finite-state languages-become rapidly intractable when the number of words of the vocabulary go beyond approximatively one or two hundred.
- 2 their model is extremely coherent ;
- 3 they can take into acount semantic constraints ;
- 4 an optimal search based on dynamic programming - DP - is possible in these machines 2.

The next section is intended to prove the last point.

III - AN OPRIMAL SEARCH SOLUTION BASED CN DP IN MULTI-LEVEL MACHINES

At the deepest level of the multi-level machine, i.e. the Olevel, the automatons are constituted only of terminal symbols which characterize the acoustic properties of the speech signal. A symbol c in a Olevel automaton is defined by its context t^0 :

$$t^0 = (n1, ||, c, ||, n2)$$

Symbol 11 means that the entry-state and the exit-state of a terminal symbol does not exist.

The Olevel automaton including symbol c in context t^0 is itself in context t^1 in the 1-ievel stage of the multi-level machine which itself is in context t^2 in the 2-level stage ... which is itself in context t^n in the n-level stage of the multi-level machine. Consequently to symbol c in context t^0 ©an be associated the list :

$$(t^n, t^{n-1}, t^{n-2}, \ldots\ldots\ldots, t^1, t^0) \qquad (3)$$

which characterizes completely such a O-level symbol in an n-level machine. Now such a symbol must be put in correspondance with the 1'th acoustical feature of the unknown sentence. This correspondance can be realized in a matching of dimensionality (n+2) where to each point

$$(t^n, t^{n-1}, \ldots\ldots\ldots\ldots, t^1, t^0, i) \qquad (4)$$

is attached the following semantic interpretation the acoustical entity is put in correspondence with a symbol in context t^0 in a machine in context t^1 in a machine in context .in a machine in context t

In order to find the best sequence of words that best matches the unknown utterance we need to define the neighbourhoods of each point of the matching space 4.

Let define $p^{k,l}$ k=1, .., 5 be the primitive projections on contexts t^l at the l'th level stage of the multi-level machine giving the k'th coordinate of these contacts.

As in the connected speech recognition problem, two kinds of neighbourhoods must be defined :

a) Inter-model neighbourhood :

this neighbourhood must be defined if $p^{1,0}$ (t^0) is not the entry-state of the O-level machine $p^{3,1}(t^1)$ in context t^1. In this case the neighbourhood of point $(t^n, t^{n-1}, t^{n-2}, \ldots, t^1, t^0, i)$ can be expressed by the following relation :

$$V\,(t^n, t^{n-1}, t^{n-2}, \ldots\ldots\ldots, t^{n-2}, t^1, t^0, i) = \bigcup_{(t'^0 \notin t^0)} (t^n, t^{n-1}, \ldots\ldots, t'^0, i-1) \qquad (5)$$

b) Inter-model neighbourhood

this neibourhood must be defined if there existes an integer l such that :

1) $p^{2,l+1}(t^{l+1}) = p^{1,l}(t^l)$ for l l n (6-1)

2) $p^{2,l-k+1}(t^{l-k+1}) = p^{1,l-k}(t^{l-k})$ for k=1,2,...,l and l l n (6-2)

If t^{n+1} designates the context of the overall machine then by definition :
$p^{2,n+1}(t^{n+1}) = p^{1,n}(t^n)$ whatever context t^n.

In this case the inter-model neighbourhood of point $(t^n, t^{n-1}, \ldots, t^1, i)$ can be expressed by the following relation :

$$V(t^n, t^{n-1}, t^l, t^0, i) = \bigcup_{(t'^l} \quad t'^l \text{ and } t'^{l-k} \text{ for } k=1,2 \ldots l$$
$$\text{s.t } p^{4,l-k+1}(t'^{l-k+1}) = p^{5,l-k}(t'^{l-k})$$
$$(t^n, t^{n-1}, t'^l, t'^{l-1}, \ldots t'^1, t'^0, i-1)$$

and the general neighbourhood of point $(t^n, t^{n-1}, \ldots, t^0, i)$ by :

$$V(t^n, t^{n-1}, \ldots, t^0, i) = V(t^n, t^{n-1}, \ldots, t^0, i) \text{ if } t^0 \notin t^0$$
$$V(t^n, t^{n-1}, \ldots, t^0, i) = V(t^n, t^{n-1}, \ldots, t^0, i) \cup (t^n, t^{n-1}, \ldots, t^0, i-1)$$
$$\text{if } t^0 \quad t^0$$

In order to express the DP recursive relation we need the following definitions :

<u>DEFINITION 1</u> : If for a particular point of the matching space $(t^n, t^{n-1}, t^{n-2}, \ldots, {}^0, i)$ with i=l relations (6) are verified for l=n and if $p^{1,n}(t^n)$ is a start of the n-level machine, then this point is called a gate-point of the matching space.

<u>END OF DEFINITION</u>

<u>DEFINITION 2</u>

- $D(t^n, t^{n-1}, t^{n-2}, \ldots, t^0, i)$ is the accumulated distance associated with the partial optimal path going through point $(t^n, t^{n-1}, t^{n-2}, \ldots, t^0, i)$.

- $d(t^n, t^{n-1}, t^{n-2}, \ldots, t^0, i)$ is a local measure of dissimilarity between the i'th acoustical feature of the unknown utterance and the acoustical representation of a O-level symbol characterized by the list of contexts $(t^n, t^{n-1}, t^{n-2}, \ldots, t^0)$.

- $d_p((t'^n, t'^{n-1}, t'^{n-2}, \ldots, t'^0, i'), (t^n, t^{n-1}, t^{n-2}, \ldots, t^0, i))$ is a local accumulated distance associated to the path joining point $(t'^n, t'^{n-1}, t'^{n-2}, \ldots, t'^0, i')$ and point $(t^n, t^{n-1}, t^{n-2}, \ldots, t^0, i)$

ED OF DEFNUTION

Owing to these definitions we are able to give the DP relations

```
IF $V(t^n,t^{n-1},\dots,t^0,i) = o$ THEN
  IF $(t^n,t^{n-1},t^{n-2},\dots,t^0,i)$ is a gate-point
    THEN $D(t^n,t^{n-1},\dots,t^0,i) = d(t^n,t^{n-1},\dots,t^0,i)$
    ELSE $D(t^n,t^{n-1},t^{n-2},\dots,t^0,i) =$
END IF
ELSE

$D(t^n,t^{n-1},t^{n-2},\dots,t^1,t^0,i) = \mathrm{MIN}_{(t'^n,t'^{n-1},\dots,t'^0,i') \in V(t^n,t^{n-1},\dots,t^0,i)}$
            $D(t'^n,t'^{n-1},\dots,t'^0,i')\ +$
            $d_p((t'^n,t'^{n-1},\dots,t'^0,i'),(t^n,t^{n-1},\dots,t^0,i))$
END IF
```

It is important to remark that as the DP relations take into acount all the levels of the multi-level machine, the solution obtained is optimal for all these levels.

From these DP recursive relations it is possible to elaborate an algorithm that can deal with any multi-level modeled language. But its explicetation is beyond the scope of this paper (3).

IV-CONCLLISION

In this paper we have presented the concepts of multi-level machines and of multi-level dynamic programming. These machines are well suited to the difficult problem of continuous speech recognition because :

1 they preserve the linguistic structure of the language ;
2 they permit the introduction of semantic constraints ;
3 they allow an optimal search based on DP.

Another interesting point of such machines and perhaps the most important one is due to the fact that the semantic links can be modified in an easy manner by the speech recognition itself. Consequently our machines have a potential hability of self-learning.

REFERENCES

1 J.E HOPCROFT, J.E ULLMAN : "Formal Languages and Their Relation to Automata", Adisson Wesley, Reading Massachussetts, 1969.

2 R.BELLMAN : "Dynamic Programming". Princeton University Press 1957.

3 J. di MARTINO : "A Multi-Level Linguistic Model for Continuous Speech Reconition", CRIN report.

4 H. NEY : "The Use of A One-Stahe Algorithm for Connected Word Recognition", IEEE Trans., Acoust., Speech, Signal Processing, VOL. ASSP-32, pp.263-271, April 1984.

A.M. COLLA, D. SCIARRA : "Automatic Generation of Linguistic, Phonetic and Acoustic Knowledhe for a Diphone-Based Continuous Speech Recognition System", in Nato Asi Series, Vol. F16, New Systems and Architectures for Automatic Speech Recognition and Synthesis, F. 16, New Systems and Architectures for Automatic Speech Recognition and Synthesis, R. DE MORI and C.Y. SUEN Editors, Sprinher-Verlag Heidelberg, 1985.

Kai-Fu Lee : "Incremental Network Generation in Template-Ba3ed Word Recognition", CMU-CS-85-181, Department of Computer Science, Carnegie-Melton University.