Integration of Rule-Based Systems and Neural Networks into Speech Recognition System

HALIMA BAHI MOKHTAR SELLAMI
Computer science department
University of Annaba
B.P.12. ANNABA
ALGERIA

Abstract:- The Artificial Neural Networks (ANN) were widely and successfully used in the automatic speech recognition (ASR) field, but many limitations inherent to their learning style remain. In an attempt to overcome these limitations, we combine in a speech recognition hybrid system the pattern processing of the ANNs and the logical inferencing of the symbolic approaches.
In particular, we are interested by the Knowledge Base Artificial Neural Network (KBANN), approach proposed initially by G. Towell [22,23]. It consists on knowledge base implemented throughout a neural network. It is an ANN where neurons have significance and the propagation of the activation represents the inferencing process in a rule-based system. In this paper, we describe a KBANN dedicated to the Arabic speech recognition.

Key words:- Artificial intelligence, speech recognition, hybrid system, neuro-symbolic integration, expert system, neural network.

1 Introduction

The AI approaches try to reproduce the natural human reasoning which incorporates several approaches of reasoning. In particular in perception problems, this allows him to recognize and to react instantly to sensory cues. This kind of hybrid intelligence has inspired AI researchers to combine multiple artificial methods and several information sources to deal with knowledge in an attempt to simulate human thought.

One of the most used integration is the neuro-symbolic one. Some researches in this area deal with the integration of expert systems and neural networks. In particular, we are interested by the knowledge base artificial neural network proposed by Towell [22,23]. We studied it and we implemented a KBANN dedicated to the Arabic speech recognition.

Our system is dedicated to the speech recognition; It is an MLP (Multi-layer Perceptron) which recognizes isolated spoken words in Arabic. With respect to Towell approach, we define first, the rules related to this field and we translated them into a neural network. Thus, we obtain an input layer which represent the acoustical level, the hidden layer the phonetic level, and the output layer, stands for the lexical one. Then the system is trained using the BackPropagation (BP) algorithm[6].

In this paper we describe our investigations throughout the expert-neural network integration, and we suggest an integration approach which we applied to the Arabic speech recognition. The paper is structured as follows, in the second section, we give a brief introduction to expert neural network and especially the KBANN. In section 3, we describe the conceptual elements of our recognizer. In section 4, the classification rules extraction principles are described. In section5, we present the final ANN topology. Finally, a discussion is done on the basis of the obtained results.

2 Expert-neural networks

2. 1. Expert systems and neural networks

An expert system consists on programs that contain knowledge base and a set of rules

that infer new facts from knowledge and from incoming data. The rules are used in the inference process to derive new facts from given ones. The strength of the expert systems is the high abstraction level, thus knowledge is declared in very comprehensive manner. The system also gives explanations for the given answers in the form of inference traces. Typical weakness are dealing with incomplete, incorrect and uncertain knowledge. Also, the system does not learn any thing by itself.

Of another side, artificial neural network is basically a dense interconnection of simple, non-linear computation elements called "neurons". It is assumed that a neuron has N inputs, labeled x_1, x_2, .., x_N, which are summed with weights w_1, w_2, …, thresholded, and linearly compressed to give the output y, defined as : $y = f(\sum w_i x_i - \Phi)$, where Φ is an internal threshold, and f is a non linearity (usually f is sigmoid function) [14,18]. They are good pattern recognizers, they are able to recognize patterns even when the data is noisy, ambiguous, distorted, or has a lot of variation.

Although, big problem in neural networks is the choice of architecture : the only way to decide on a certain architecture is by trial-and-error. Another weakness of neural networks is the lack of explanation.

2. 2 Expert neural networks

The researchers tried to overcome expert systems and neural networks limitations by creating hybrid systems. Various classification schemes of hybrid systems have been proposed [13,17,21,24] As a brief introduction, we present a simplified taxonomy, where such systems are grouped on two categories : transformational and coupled models.

a) In the transformational models (transnational as [13]), the expert system could be transformed to a neural network or the neural network could be transformed to an expert system.
b) In the coupled models, the application is constituted of separated two components, that can exchange knowledge. Neural network can be used like component of pre-treatment for the expert system. The expert system can prepare data for neural network and can contribute to the determination of the network configuration. The KBANN can be considered as a coupled system.

2. 3. KBANN

KBANN starts out with knowledge base of a set of rules. This domain theory has to be translated into a neural network. The KBANN system uses a seven step algorithm to construct a network.

1. Rewriting : the rules are transformed into Horn clauses. Rules with same consequence are rewritten in the following form : A:-B,C,D. A:-E,F. To : A:-A'. A:-A''. A':-B,C,D. A'':-E,F.
2. Mapping : the rules are then organized into a neural network. The weight values are chosen in such a way that the activation emulates an AND-function.

   a) Translation of conjunctive rule into KBAN-net : If we have the following rule : A:-B,C,D,not(E). It will be translated to a network like the one below (fig.1). All links corresponding to positive antecedents are set to ω, and all links corresponding to negative antecedents are set to $-\omega$. Bias θ corresponding to the consequent of the rule is set to $(P-1/2)\omega$. The suggested value of ω is 4.

   b) Translation of disjunctions : If we have the following rules : A:-B. A:-C. A:-D. A:-D. This is translated to a network as in fig.2. The bias in this case is set to $-\omega/2$, so that, any of the input neurons should be able to activate the output one.

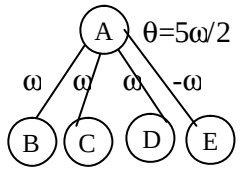

fig.1. conjonction net

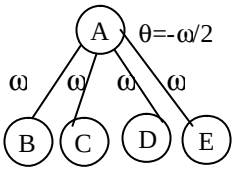

fig.2. disjonction net

3. Numbering : each node in the network is assigned a number according to its "level".
4. Adding hidden units : new units are placed in the network to facilitate

learning of features that were not present in the initial rules. This step is optional, because the given rules often have enough expressive power to make the learning of new rules possible.

5. Adding input units : the domain expert can identify certain features that were not caught in a rule.
6. Adding links : links with weight zero are added to the network. Each node at level n-1 is connected to all nodes with level n.
7. Perturbing : in order to ovoid problems caused by symmetry a small random value is added to each weight in the network.

Finally, Towell lists the correspondence between a domain theory and a neural network as :

Domain Theory	Neural Network
Final Conclusions	Output Units
Intermediate Conclusions	Hidden Units
Supporting Facts	Input Units
Antecedents of a rule	Highly weighted links

In KBANN all functions are differentiable, so, backpropagation is applied to all layers.

3 Neural networks in ASR

The neural networks were widely and successfully used in pattern recognition, in speech recognition many improvements were made to increase their recognition rate[12]. As it is commonly agree, the difficulty in using neural networks is how to configure the neural network, and what are initial weights of links between neurons. Since, we believe that, the speech recognition application we are considering could be expressed as rules, we tried to follow the Towell's approach to set the network topology.

3.1 Propositions

In the Towell's approach only propositional logic is considered, so our "target" vocabulary should be composed by propositions. At first time, these propositions are extracted from the domain theory, and they will represent some neurons in the network. We defined them as follows : The propositions related to the output layer are the vocabulary words, in our case : the ten Arabic digits. They were formed using the syllables, so we add, as many propositions as there are syllables implied by the ten digits pronunciations. In parallel, these syllables are recognized by some acoustic characteristics, these characteristics represent the propositions related to the input layer.

3. 2 Horn Clauses

As said already, the words are formed using syllables, so the corresponding rules are : Wordi:-syl_1, syl_2, etc. The rules related to the obtention of syllables are of the form : syl_i :- c_j, c_k,etc. We notice that, we have only positive literals. C_i, are acoustic classes obtained by using the vector quantization (VQ) techniques.

Overall, our approach could be decomposed in two stages : the first one consists on the collect of the acoustic and phonetic knowledge implied by the network conception. Then, when the network is built, it must be trained and the recognition tests will be done.

4. Classification rules

Firstly, we are going to determine the acoustical classes related to our vocabulary. The features extraction stage will provide a set of acoustical vectors, in order to get discrete values, we transform them using the VQ.

4. 1 Features extraction

The signal of the spoken word is sampled at a rate of 11025Hz. Then, all background before and after the word is eliminated. After that, in the first stage, the word is segmented into syllables and for every obtained wave file, we proceed to the features extraction. In the second stage (after the MLP conception) the signal of the whole word is analysed similarly. The steps we followed are [1,2,19] :

- Preemphasis : The sampled signal is processed by a first-order digital filter in order to spectrally flatten the signal.

- Blocking into frames : Sections of N consecutives samples are blocked into a single frame (N = 512 samples of signal). Frames are spaced M samples (M = 256).
- Frame windowing : Each frame is multiplied by a N-sample Hamming window.
- Autocorrelation analysis : Each windowed set is autocorrelated to give a set of (p+1) coefficients, where p is the order of the LPC analysis.

$$Rt(m)=\sum_{N=1}^{N-M} xt(n)\times xt(n+m), 0\le m\le p, p=8$$

- LPC/Cepstral analysis : For each frame, a vector of LPC coefficients is computed from the autocorrelation vector using the Levinson method [8]. An LPC derived cepstral vector is computed with q coefficients, with q>p, we use q=12.
- Cepstral weighting : The cepstral vector of q component $c_t(m)$ at time frame t is weighted by a window $w_c(m)$ of the form:

$$W_c(m)=1-q/2\times\sin(p\times m/q)$$
$$x_t(m)=c_t(m)\times W_c(m), 1\le m\le q$$

- Delta cepstrum : The time derivative of the sequence of weighted cepstral vectors is approximated by a first-order orthogonal polynomial over a finite length window of (2K+1) frames, centred around the current vector (we use k = 2, hence a 5 frame window is used). The cepstral derivative (or the delta cepstrum) is computed as:

$$\Delta c_l(m)=[\sum_{k=-K}^{K} k c_{t-K}(m)]\times G, G=0.375; 1\le m\le q$$

The acoustical vector is the concatenation of the weighted cepstral vector, and the corresponding weighted delta cepstrum vector, i.e., $V_t=\{c_t(m), \Delta c_t(m)\}$;$1\le m\le q$. To each window of the signal will correspond a numerical vector of 24 coefficients.

4. 2. Vector quantization (V-Q)

Given a training set of continuous observation vectors, the V-Q partitions the training vectors into M disjoint regions (M is the size of the codebook), and represents each set by a single vector v_m, which is generally the centroïd of the training set assigned to the m^{th} region.

We consider all acoustical vectors we obtain during the training stage, we regroup them into disjoint classes using the LBG algorithm, a variant of the k-means [15]. The prototypes we obtain represent acoustical frames, and will constitute the entries of the code-book.

At the recognition phase, the vector quantizer compares each acoustical vector v_j of the signal to stored vectors c_i, that represent the code-words, and v_j is coded by the vector c_b that best represents v_j according to some distortion measure d. $d(v_j, c_b) = \min(d(v_j, c_i))$.

4. 3. Syllable : the decision unit

The Arabic speech has the particularity to present few vowels, few consonants and a regular structure of syllables [1,9,11]. Syllables could also be easily processed and have well defined linguistic statute, especially in the phonetic level where they represent suitable unit for the lexical access. These elements have motivated our choice to consider the syllable for modelling the phonetic level. Another element sustained this choice, which is, given a set of Arabic syllables in the nearly totality of cases only a single word could be formed.

So, the prototypes, we define in the earlier stage, will characterize a syllable if they exist or not in a given signal.

In the following, we present the five possible patterns of Arabic syllable presented in [9], In their representation C, stands for all consonants, V for short vowels and VV for long vowels. The first four patterns occur initially, medially and finally. The fifth pattern, occurs only finally or in isolation.

1. **CV** eg. /bi/ (with)
2. **CVC** eg. /sin/ (tooth)
3. **CVV** eg. /maa/ (not)
4. **CVVC** eg. /baab/ (door)
5. **CVCC** eg. /sifr/ (zero)

4. 4. Classification rules

The network architecture reproduces two kinds of rules, the first one represents the relationship between the input and the hidden layer and the second between the hidden and the output layer.

The first kind of rules have the following form : **If** conjunction (classes) **then** syllable. They explain connections between the input and the hidden nodes. The input nodes are classes issued form the V-Q stage, we decided to keep 32 classes. Every one of these classes represent an acoustical characteristic of the signal. Since, we are interested with the ten Arabic digits an example of these rules is :

If C_1 **and** C_2 **and** C_7 **and** C_8 **and** C_{19} **and** C_{20} **and** C_{28} **and** C_{30} **then** γam

The second kind of rules have the form : **If** conjunction (syllables) **then** word. They represent the relation between hidden and output nodes. Here is an example :

If sifr **then** sifr. Here the digit "sifr" (zero) is composed by 1 syllable.

5 From rules to network

The neural network is a multilayer perceptron, we built it while being careful to respect the Towell procedure. It has the following characteristics: The input layer, contains thirteen (32) neurons representing the whole acoustic classes. A signal in entry of the system is analyzed, then transformed to a symbolic chain by the vector quantizer. Each symbol is the index corresponding to the prototype of the vector in code - book. Every entry of the network is going to receive a binary value (1 or 0) according to the existence or no of the corresponding characteristic in the signal.

The output layer, contains ten (10) neurons representing the words of the vocabulary, in our example we have the ten Arabic digits.

The first layer from the hidden ones, contains twenty nine (29) neurons corresponding to the syllables related to the pronunciations of our vocabulary. Since, one word has more than one pronunciation,(see 2.3.), we added lot of nodes representing, we interpret as the various pronunciations of the ten digits. These neurons stand in second hidden layer. Once the network topology defined, we train it by using the BP algorithm.

6 Related works

A similar work was performed, while following the approach of S. I. Gallant[3,4,5]. Gallant, was the first to describe a system combining the domain expert knowledge with neural training [10]. The system starts with dependency information from which it builds a structured neural network with only feedforward connections without cycle. Contrary to KBANN approach, we consider only specified connections, so that, connections not specified could never be discovered. To train the network, Gallant suggests the use of a relevant algorithm called: Pocket algorithm. The Pocket algorithm is a modification of the perceptron learning.

7 Discussion

We perform some tests to evaluate the KBANN performances comparatively to other approaches in Arabic speech recognition. We consider a training corpus constituted by three speakers; each of them uttered three times the ten digits. The test corpus, comprise four speakers each of them uttered twice the ten digits. In the table bellow, we mention results; we have obtained with various implementations.

a. with a classical MLP trained with BP algorithm.
b. with Hidden Markov Models [2].
c. with CES, 17 hidden neurons [3,4]
d. with CES, 29 hidden neurons [5]
e. with KBANN, 29 hidden neurons.

(a)	(b)	(c)	(d)	(e)
94	95	95.71	97.86	98

We also notice that, when using domain knowledge, the network train faster and generalize better than the classical ones.

Another aspect appears when we assume a knowledge-based approach relying on the recognition phase. In the recognition phase, a word could be recognized and well categorizes or recognized and badly categorized; this addresses the question of

the system reliability, so that the explanation aspect becomes too much important. This aspect is absent in the connectionist models but is effectively present in rule-based systems.

Overall, the neural expert models are a promising trend in resolution of perception problems, since this category of problems involve both neuronal models and symbolic reasoning.

References

[1] H. Bahi, M. Sellami, « An acoustical based approach for arabic syllables recognition », workshop on software for the arabic language, Beirut, Lebanon, June 2001

[2] H. Bahi, M. Sellami, "Combination of vector quantization and hidden Markov models for Arabic speech recognition", ACS/IEEE Proceedings of AICCSA'01, Beirut, Lebanon, June 2001, pp96-100.

[3] H. Bahi, M. Sellami, "Hybrid approach for speech recognition", IAPR Proceedings of ICISP'03-Agadir, Marocco, June 2003.

[4] H. Bahi, M. Sellami, "Hybrid approach for Arabic speech recognition", ACS/IEEE Proceedings of AICCSA'03-Tunis,Tunisia, Jul. 2003.

[5] H. Bahi, M. Sellami, « Système expert connexioniste pour la reconnaissance de la parole », to appear in RFIA proceedings, Toulouse, France, 28-29 Jan. 2004.

[6] C.M. Bishop " Neural networks for pattern recognition ", Clarendon Press, Oxford, 1995.

[7] O. Boz, « knowledge integration and rule extraction in neural networks », university of Lehigh, 1995.

[8] Calliope, " La parole et son traitement automatique ",Masson, 1989

[9] S. H. El-ani, " Arabic phonology, An acoustical and physiological investigation", Indiana university, 1989

[10] S. I. Gallant, "Connectionist Expert Systems", Communications of the ACM, Vol. 31, N°2, Feb.1988, pp:152-169

[11] M. Harkat, " Les sons et la phonologie ", ed. Dar el afaq, Algeries 1993

[12] J. P. Haton, "Les modèles neuronaux et hybrides en reconnaissance automatique de la parole : état des recherches ».Rapport of CRIN/INRIA.

[13] M. Hilario, "An overview of strategies for neurosymbolic Integration"

[14] J. F. Jodouin, « Les réseaux neuromémitiques : modèles et applications », éd. Hermès, Paris, 1994

[15] Y. Linde, A. Buzo, R. M. Gray, "An algorithm for vector quantizer design", IEEE transactions on Computer, N°36, pp : 84-95, 1980

[16] R. Louie, «hybrid Intelligent Systems Integration into Complex Multi-Source Information systems », Master Thesis, MIT, Aug. 1999.

[17] L, R. Medsker, "Hybrid neural network and expert systems", Kluwer academic publishers, 1995.

[18] N. Morgan, H. A. Bourlard, "Neural networks for statistical recognition of continuous speech ", Proceeding of the IEEE, Vol. 83, N°5, may 1995

[19] L. Rabiner, B. Hwang, "Fundamentals of speech recognition", Prentice Hall, 1993.

[20] J. Sima, « Neural expert system », journal neural networks, Vol. 8, Number 2, pp : 261-271, 1995.

[21] J. Sima, "Review of integration strategies in neural hybrid systems", citesser.nec.nj.com/

[22] G. G. Towell, «Symbolic knowledge and neural networks : Insertion, Refinement and extraction », Thesis of doctorat, University of Wisconsin, Madison, 1991.

[23] G. G. Towell, J. W. Shavlik, "Knowledge-based Artificial Neural Networks", Artificial Intelligence 70, 1994, pp:119-165.

[24] S. Wermter, R. Sun, "Hybrid Neural Systems", Springer, New York, Jan. 2000.