Hybrid hierarchical steps and learning by abstraction for an emotion recognition system *

BRUNO APOLLONI, CHRISTOS OROVAS, GIORGIO PALMAS
Computer Science Department, University of Milan
Via Comelico 39, 20 135, ITALY
apolloni@dsi.unimi.it, orovas@media.dsi.unimi.it, palmas@laren.usr.dsi.unimi.it

Abstract: In this paper the hierarchical approach in learning and its application in building a hybrid (symbolic - subsymbolic) system for emotion recognition is examined. Using the hierarchical approach, instead of creating a solid and direct mapping from a set of inputs to the set of the corresponding outputs, the process is divided into a number of stages where the goal is to construct an abstract description of the concept to be learned. This principle is followed by PAC meditation which is a probably approximate correct (PAC) learning paradigm for Boolean formulas. An extension of this system in order to allow the expression of more complex hypotheses is to incorporate subsymbolic processing at various stages.

Keywords: PAC learning, PAC meditation, hybrid systems, abstraction learning, emotion recognition

1 Introduction

One way to define learning is as the automatic construction of the mapping from a set of stimuli to the set of actions to be taken by an agent given these stimuli. According to the nature of the data being processed, the methods which are followed and the form of representation of the knowledge which is necessary in order to perform the mapping we can distinguish the statistical learning (or inferencing), the subsymbolic learning (in artificial neural networks) and the symbolic learning (in the sense of rule based systems).

In an effort to examine the feasibility of learning problems when approached under a statistical oriented prism, the probably approximate correct (PAC) learning notion was introduced [1, 2]. In that, learnability is defined by whether or not a sufficiently close approximation of a target concept can be derived by a polynomial time learning algorithm. Learnability of concepts from the Boolean space was initially examined [1]. The model was then extended for concepts from a real valued instance space [3] and a structural measure of the class of concepts, the Vapnic-Chervonenkis (VC) dimension [4, 3], became a basic parameter connected with the learnability of a class and the calculation of the sample sizes required. Defining a sentinel function and working with the frontiers of the concepts to be learned was then suggested in [5]. Extending the notion of sentinels by including both the inner and the outer frontiers of the concepts and by following a multistage approach where starting from initial observations about a concept the level of abstraction is gradually increased until a high level description is achieved was then reported in [6]. That was known as the PAC meditation model and a learning paradigm for Boolean formulas was given. As it was shown in [6], following this approach the solution of learning problems that had been reported as NP-hard (e.g. learning k-term-DNF functions [7]) was proved NP-easy [8].

The merits of learning by abstraction and constructing hierarchical descriptions of concepts were also recently reported in an example for image understanding [9]. In that, a cellular communication protocol is adopted and is augmented by the use of symbolic descriptions and neural associative processing of the rules. This system is also a paradigm of a hybrid architecture where symbolic and subsymbolic processing coexist and supplement each other [10]. The study of such systems and their application for a challenging HCI problem, that of emotion recognition, is the subject of the PHYSTA research project.

In this paper, after a brief presentation of PAC

* *Research within the framework of TMR PHYSTA project. Research contract FMRX-CT97-0098 (DG12-BDCN)*

learning, PAC meditation and the issues regarding the problem domain, the initial framework for a two way communicating hybrid system combining subsymbolic processing at the initial stages and the creation of hypotheses using the PAC meditation model at the later stages is given.

2 PAC learning

As mentioned earlier, PAC learning is referring to the feasibility of learning problems as far as computational complexity and performance is concerned. Performance in this case is related with the level in which the constructed hypothesis approximates the concept to be learned. Thus, having X as the set of all possible outcomes of a source of random data, we define a *concept* c to be any subset of X and a *concept class* any set of concepts. Points are drawn from X with an arbitrary probability distribution P and using a *characteristic function* χ_c ($\chi_c : X \rightarrow \{1, 0\}$) they can be labelled as belonging to a concept c or not. A *labelled random sample* ξ_m^c for each concept c is then defined as the set of the m pairs $\{(\xi_i, \chi_c(\xi_i)), \quad i = 1 \ldots m\}$ where ξ_i are randomly chosen from X with probability $P(\xi_i)$. A *learning function* for C is a procedure that given a large enough labelled random sample for any concept c of C generates a hypothesis h_c which is a good approximation of c. In order to define what is meant by a good approximation we need to define the *error* between a concept and a hypothesis. This is defined as the probability, according to P, of the symmetric difference between the concept c and the hypothesis h_c (i.e. $P([(c \cup h_c) - (c \cap h_c)])$). In other words, the error is the probability that h_c and c will disagree on an instance randomly drawn from X. A good approximation is then the one for which the probability of the error being more than ϵ is less than δ for small ϵ and δ. A class C of concepts will be called *PAC learnable* if there exists a learning function which for any probability distribution P can run in time polynomial to $1/\epsilon$, $1/\delta$ and the dimensionality of X and can produce for each concept of the class a hypothesis which is a good approximation of it according to P, ϵ and δ.

The case of three categories of Boolean functions (namely, k-CNF, k-term-DNF and μ-expressions[1]) was initially examined in the introductory paper by Valiant [1]. The use of the routines EXAMPLE and ORACLE from the learning algorithms was considered in that case. A call at EXAMPLE would return a positive example and ORACLE was a routine which in its simplest form could tell if an instance from $\{0, 1\}^n$ belonged to the concept or not. According to [1], the first category was learnable from positive examples only while the next two categories needed calls to complex versions of ORACLE as well. However, there were classes of Boolean functions that proved not PAC learnable without ORACLEs. For example the k-term-DNF Boolean functions for $k \geq 2$ [7].

A structural measure, the Vapnic-Chervonenkis dimension[2] (d_{VC}) [4, 3] was related to the PAC learnability of a concept class in [3] where concept classes defined by regions in the Euclidean space, E^n, are also considered. According to [3], a concept class is PAC learnable iff the VC dimension of C is finite. Additionally, the sample sizes required for learning the concepts are defined in relation with ϵ, δ and the VC dimension. The smallest sample size that allows this is called the *sample complexity* of the learning function.

A different approach to define the sample complexity was suggested in [5]. That was derived from the observation that the points in the sample play the role of 'sentinelling' the target concept. Moreover, it is possible that a subset only of these points would suffice in order to represent the same concept. The new index of the sample complexity for C was called the *detail* of C and was defined as the supremum of the minimum numbers of sentinel points required for each concept in C. Based on this notion, the sentry function which would return the set of points sentinelling a concept from outside was defined. Combining the latter ideas with a multilevel approach in order to build hypotheses for the target concepts is the PAC meditation model [6].

[1] A k-CNF function is a conjuction of clauses with at most k literals in each clause, a k-term-DNF function is a disjunction of at most k monomials and an μ-expression is an arbitrary Boolean expression containing each variable at most once.

[2] Simply stated, the VC dimension of a concept class $C \subseteq 2^X$ is the maximum number d of instances from X that can be labelled as positive and negative examples in all 2^d possible ways through concepts from C.

2.1 PAC meditation

The basic characteristic in PAC meditation is that the learning process is divided into a number of stages. These stages are the consecutive steps in a two sided converging approximation procedure where at each stage the level of abstraction is increased. Thus, starting from initial observations about the concept to be learned (i.e. from the labelled random sample) a number of partial consistent hypotheses are created. That is, each hypothesis is consistent with a part of the positive examples and all negative examples or vice versa. The basic point at this stage is that the union of the hypotheses coming from positive subsets and the intersection of the hypotheses coming from the negative subsets form two inner and outer frontiers respectively. These two frontiers delimit the interstice where the suitable consistent hypothesis can be found (see fig. 1a).

At further abstraction levels, the partial consistent hypotheses of the immediately preceding level play the role of labelled examples. This is depicted in figure 1b. Thus, the sampled data are substituted by formulas of these data and the positive/negative labels by flags indicating whether the formulas belong to the inner or the outer frontier. New links can then be stated between the frontier formulas. Moreover, a redefinition of the interstice for the target concept by means of extending the inner and 'shrinking' the outer frontier can take place. These two steps (i.e. new links and redefinition of interstice) are called *abstraction* and *reduction* respectively.

The learning model in PAC meditation can be thought of as a 'bridge' between inductive and deductive learning [6]; at the first level atomic formulas are inductively learned from examples and then these formulas are processed with special deductive tools. Moreover, the model provides a special featured boosting of hypotheses which could be called syntactical when compared with other methods which could be called statistical [11]. The basic idea of boosting is to start from weak hypotheses suitable only in different regions of the instance space and combine them in order to produce a strong hypothesis. The way of combing the hypotheses can have a numerical nature, in which case is reminiscent of methods for combining multiple experts [12], or, as in the example of PAC meditation, can have a more delicate form. In the latter, the structural complexity of the hypotheses grows with the level and denotes an incremental embedding of new symbolic knowledge. Thus, the derived hypothesis is not only consistent and representative of the target concept but also expresses more intricate relations among the observed data, the stimuli, and their high level interpretation.

As mentioned in the previous section, PAC meditation uses the notion of sentry functions in its learning mechanism. These are now extended in order to include the inner and the outer sentry function which given a hypothesis return the internal and external pivots, respectively, around which to draw that hypothesis. Estimations of the sample complexity is also allowed by using these functions.

In the operation of the model special care has been taken so that the process is *without information waste (wiw)*. An information waste occurs when relying to overdetailed hypotheses which require the model to focus on sentry points that in later stages will be proved useless for drawing the final hypothesis.

In the example for Boolean formulas which is given in [6], a PAC meditation algorithm which is called *systolic meditation* is presented. In that, families of *hyper_L_clauses*[3] and *hyper_L_monomials* are constructed starting from a labelled random sample and the procedure stops when the planned hierarchical level has been reached. Thus, a series of abstraction and reductions steps is followed and when the proper level is reached *synthesis* takes place. The latter is the final step and it is the one in which the final hypothesis is drawn. Depending on the desired form for the final hypothesis we can either group r of the possible hyper_L_monomials or r of the possible hyper_L_clauses representing the inner and the outer frontiers of our target concept respectively. The basic criterion which has to followed in each case is that of the consistency of the drawn hypothesis with the inner and outer frontiers.

[3] A hyper_L_clause is a disjunction of hyper_L-1_monomials and vice versa. A hyper_0_clause is the usual clause (disjunction of literals) and a hyper_0_monomial is the usual monomial (conjunction of literals).

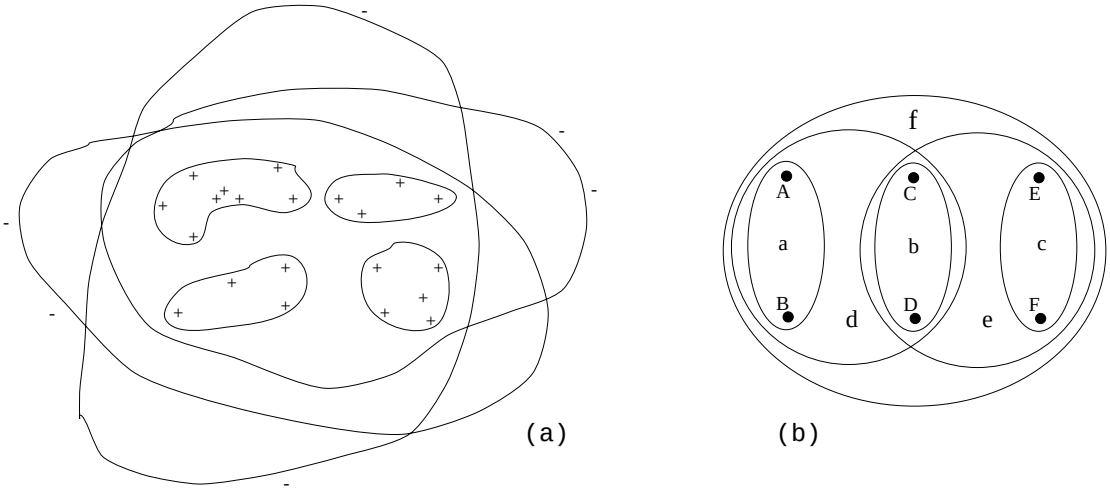

Figure 1: (a) The partial consistent hypotheses from negative and positive examples. (b) Sentry points at various abstraction levels.

The application of the PAC meditation model for learning concepts within a more diverse and extended class of concepts is in the scope of the current research. The new instance space includes examples and observations derived after the proper preprocessing of data which are necessary for emotion recognition.

3 Emotion recognition

The development of an artificial emotion decoding system is of great interest and it is a challenging application in the field of human-computer interaction (HCI). Within the framework of the PHYSTA project the problem is approached at different levels. At the first level there is the raw input signal, the stimuli, which has to be preprocessed in order to extract these features which are necessary for the further stages of processing. Vision (static and dynamic images of faces) and acoustic signals (speech) are the input data to this level. Features from frontal and profile views of faces, optical flow and speech signals must be extracted at this stage [13]

At the intermediate level there is the connection of the features with the emotional status which is expressed in the video and acoustic sequence. This can be interpreted as 'which values of a feature are characteristic of the current emotional status ?' or 'what is the interrelation of the values of the features at a given emotional status ?' etc. Questions of this kind should be handled and be the subject of symbolic manipulation when properly augmented with subsymbolic processing in order to increase flexibility.

At the highest level there exists the description and the classification of the emotions themselves. However, a difficulty arises at this point due to the lack of a standard 'vocabulary' and unified theory for this description. The approach followed by PHYSTA is presented in [14] and suggests the construction of a basic emotion vocabulary and the description of the basic emotion terms by means of (i) a two dimensional space where emotions are organized according to how positive or negative they are and the energy level connected with the person experiencing the emotion, and, (ii) an emotion schema which provides additional discriminant dimensions and deals with actions, objects and situations related with the emotion terms.

The task of the hybrid architecture which will combine the PAC meditation model and subsymbolic processing is to form the suitable mappings from the features space to the emotions space. The initial framework and related issues are presented next.

4 The subsymbolic framework

In the PAC meditation learning paradigm for Boolean formulas which was briefly presented earlier we can distinguish the following stages:

1. A set of monomials and clauses are extracted from the labelled samples.

2. Grouping of the frontiers using the set union and intersection functions.

3. Redefinition of the frontiers and derivation of higher level pivot points.

4. Final grouping of the frontiers in order to draw a consistent hypothesis.

As we saw, this model can create a set of hypotheses which are consistent and approximate the target concept with a defined accuracy. Subsymbolic extensions at the operation of the system can provide an adaptable interface with the external world and also a flexible way to define the strategy followed to build the hypotheses. These extensions are based at the suggestions given in [6].

At first level we need to define the input variables, the 0-level literals in the system. These literals should be capable of describing, at the lowest level, emotion related characteristics, e.g. shape of mouth and eyebrows, size of pupil, pitch and loadness changes of voice, etc. In other words, they should describe the elementary facts about the outside world. In a sense, this is a first step in reducing the dimensionality of the input domain. This can be achieved by a relatively primitive feature to symbol mapping following the initial feature extraction from the input signals [13]. One or more MLP networks operating in parallel can be used for this. The binary values that we need to associate with the symbols in order to start the symbolic processing with PAC meditation can then be derived by thresholding the outputs of the initial feature to symbol mapping.

At the next stage, the way of grouping and combining the initial hypotheses in order to gain higher abstraction levels can be supplemented by incorporating a 'weighting' notion. Thus, higher level hypotheses are build around symbols with the higher probability of not disappearing (i.e. not prove to be useless) at the next stages. The weights in this procedure can be determined in an adaptive way according to the 'correctness' of the derived higher level hypothesis, or be defined as to express a variety of external factors (e.g. cultural, educational, nationality and other similar parameters regarding the user). Of course, a combination of both may exist, e.g. the starting values of the weights carry information regarding these parameters and then a fine-tuning process for each user can take place. Another way for combining the hypotheses at each stage is by using equivalence relations between them. These relations can also be based in the same external factors or set in an adaptive manner or both. The basic principle in the process of forming the higher level hypotheses is that no information waste occurs and hypotheses are as generic and representative as possible. Thus, characteristics that prove equivalent would not be preferred as the components of a new hypothesis due to the potential loss of information compared to the case that these characteristics were distributed in different hypotheses.

A block diagram of the operation of the system as described above is given in figure 2. In that, we see that the audio and video signal is initially subjected to a signal to feature mapping. The extracted features are then combined in order to define an initial description of what is happening using low level symbolic terms. Applying a threshold function we can then derive a more discrete description and use it in order to build the hypotheses regarding the ways that these basic symbols are related in each emotional state. It must be noted that up to the initial symbols' level the system can be 'pre-trained'. Thus, the activation levels of the output nodes of this stage will be related to the 'elementary facts' they represent.

The issues that have to be considered at this point, especially when a correct set of hypotheses cannot be build, are the following: (i) the examples that we use in order to build the hypotheses might not be of the same 'sharpness' level. That is, the activation levels of the corresponding units may not be very close to 0 or 1 but somewhere in the middle, (ii) the activation space of these units might not be divided in two areas of 0 and 1 only but more than two areas, not necessarily consecutive, may correspond to an interpretation as 0 or 1. This could be an indication that more symbols to represent the basic facts about the external environment are needed. Alternatively, we can allow these symbols to self-adjust their meaning in

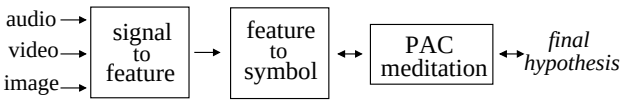

Figure 2: Block diagram of the system.

accordance to the concept that we are trying to learn, (iii) the strategy that we are following in order to build the hypotheses might not be correct and a fine-tuning process, as mentioned earlier, must take place.

These issues and their interrelations have to be further investigated at this stage. The level at which we can rely on a specific example, the existence of 'grey areas' in the concepts space and the priority of the error propagation when a hypothesis is not correct are the current subjects of examination.

5 Summary

A brief presentation of the PAC learning notion, the PAC meditation model and the issues regarding the construction of a hybrid system which can be applied to the emotion recognition problem were given in this paper. As it was shown in [6], a hierarchical approach where simple hypotheses are first drawn from the sample data and then they are used in order to define more abstract hypotheses can help reducing the complexity of the learning problem. Combining this approach with a subsymbolic processing stage in order to create a low dimensional and adaptable interpretation of elementary facts about the external environment and allowing a flexibility at the strategies followed in order to build hypotheses at a symbolic level can then be the basis for an effective hybrid system where transitions from the subsymbolic to the symbolic level and backwards in order to fine-tune both stages exists.

References:

[1] L.G.Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

[2] D. Haussler. Probably approximate correct learning. In *Proc. of the 8th Nat. Conf. on AI*, pages 1101–1108, 1990.

[3] A.Blumer et al. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM*, 36:929–965, 1989.

[4] V.N. Vapnik. *Estimation of dependencies based on empirical data.* Springer, 1982.

[5] B. Apolloni and S. Chiaravalli. PAC learning of concept classes through the boundaries of their items. *Theoretical Computer Science*, 172:91–120, 1997.

[6] B. Apolloni, F. Baraghini, and G. Palmas. PAC meditation on boolean formulas. Technical report, Dept. of Computer Science, University of Milan, 1998.

[7] L. Pitt and L.G.Valiant. Computational limitations on learning from examples. *Communications of the ACM*, 35(4):965–984, 1988.

[8] M.R. Garey and D.S. Johnson. *Computers and intractability: a guide to the theory of NP-completeness.* Freeman, 1979.

[9] C. Orovas and J. Austin. A cellular system for pattern recognition using associative neural networks. In *5th IEEE International Workshop on Cellular Neural Networks and their Applications*, pages 143–148, London, April 1998.

[10] R. Sun and L.A. Bookman, editors. *Computational architectures integrating neural and symbolic processing.* Kluwer Academic Publishers, 1995.

[11] R.E. Schapire. The strength of week learnability. *Machine Learning*, 5:190–227, 1990.

[12] R.A Jacobs. Methods for combining experts' probability assessments. *Neural Computation*, 7:867–888, 1995.

[13] Development of feature representations from emotionally coded facial signals and speech. Report for PHYSTA, *http://www.image.ntua.gr/physta*,1999.

[14] R. Cowie et al. What a neural net needs to know about emotion words. *(in the same session)*, 1999.