Algorithms of hierarchical mixture of opinions of experts in problems of synthesis of information management systems city development

Irina Pocebneva^{1,*}, *Vadim Belousov*¹, *Irina Fateeva*¹, *Vitaly Lukinov*² and *Tatyana Folomeeva*³

¹Voronezh state technical university, 20 anniversaries of October, 84, 394006, Voronezh, Russia

²Moscow State University of Civil Engineering, Yaroslavskoe sh. 26, Moscow, 129337, Russia

³Moscow state University of technology and management. K.G. Razumovsky, Zemlyanoy Val street, 73, Moscow, 109004, Russia

Abstract. In the present article we will consider a class of associative machines with dynamic structure where the entrance signal exerts direct impact on the mechanism of association of output signals of experts. At the same time we are interested in such group of expert decisions at which separate expert responses unite not linearly through hierarchically organized lock networks. Hierarchical mixture of opinions of experts, along with simple mixture are examples of modular networks: neural network of a module if the calculations executed by it can be distributed on several subsystems processing different entrance signals and not crossed in the work. Output signals of these subsystems unite the integrative module which exit does not possess feedback with subsystems. In fact, the integrative module makes the decision as output signals of subsystems are grouped in the general output signal of system, and identifies what examples are samples for training of concrete modules. The most general definition of modular neural network: any set of algorithms of data processing, including algorithms of the artificial neural networks grouped for the solution of some uniform task. Automatically determine the class of associative machines with dynamic structure where the entrance signal exerts direct impact on the mechanism of association of output signals of experts, at the same time group of expert decisions at which separate expert responses unite not linearly through hierarchically organized lock networks is considered.

1 Introduction

Many phenomena of modern reality in various areas (technical researches, economy, sociology, medicine and other) assume removal of some common decision on the basis of a

*Corresponding author: ipocebneva@vgasu.vrn.ru

set of small and simpler tasks. In other words, the complex computing challenge can be simplified due to splitting entrance space by experts into a set of subspaces. A similar expert combination call the associative car [1]. Her idea appeared in the mid-sixties the last century and consists in integration of knowledge accumulated by a great number of experts into the common decision which is priority over separately taken expert opinion. The called property is especially valuable to such difficult systems where the inaccuracy and subjectivity of the concrete decision can be critical for all structure.

2 Problem definition

Hierarchical mixture of opinions of experts, along with simple mixture are examples of modular networks which formalized definition is given in [2]: neural network modular if the calculations executed by it can be distributed on several subsystems processing different entrance signals and not crossed in the work. Output signals of these subsystems unite the integrative module which exit does not possess feedback with subsystems. In fact, the integrative module makes the decision as output signals of subsystems are grouped in the general output signal of system, and identifies what examples are samples for training of concrete modules. The most general definition of modular neural network is given in [3]: any set of algorithms of data processing, including algorithms of the artificial neural networks grouped for the solution of some uniform task.

In Fig. 1 the model of hierarchical mixture of opinions of four experts is presented (hierarchical mixture of experts - HME). It represents expansion of model of simple mixture of opinions of experts (mixture of experts - ME).

Fig. 1. Model of hierarchical mixture of opinions of experts (HME) on the example of two-level hierarchy.

It is obvious that the architecture of the HME model reminds a tree in which branches are networks of locks, and leaves are experts. The entrance space breaks into the enclosed

subspaces, the arriving information is grouped and redistributed between experts under control of networks of the locks structured hierarchically. From Fig. 1 it is visible that the HME model includes two layers of networks of a lock (two levels of hierarchy). If to apply the principle "share and dominate" [4], then it is possible to represent the HME model with a large number of levels.

There are two main ways of the description of hierarchical model of mixture of opinions of experts [5]. They are presented in Fig. 2.

It is obvious that the second way of the description of HME model is preferable: if to consider the HME model from a position of a probabilistic basis for formation of a tree of decisions, then it will be presented possible to calculate similarity for any set of data.

Now we will pass directly to the choice of the HME model, that is determination of quantity and composition of knots of the decision. As the HME model has similarity to a standard tree of decisions (for example, a tree of classification and regression - classification and regression tree - CART), we will reduce this model to start of CART on data of training of an algorithm and to acceptance of a tree which we received as initial at a stage of initialization [5].

Fig. 2. Formal ways of the description of HME model.

In Fig. 3 the example of a tree of CART where the entrance space of X breaks consistently a series of binary divisions on, so-called, terminal knots is presented.

If we compare Fig. 1 and 3, then we will be able to mark out some general properties of the CART and HME models: to networks of locks in HME the rule of the choice of splitting in CART model branches (intermediate knots) plays a similar role; to networks of experts in HME terminal knots in CART are similar.

Thus, if to apply the CART model to a problem of regression or classification, then advantages of the discrete nature of this model will allow to carry out effectively search among possible alternative trees, and the choice of such tree as initial at a stage of initialization will provide a continuous probabilistic basis in the HME model and will receive the improved soft assessment in a desirable response.

Fig. 3. Tree of decisions with a series of binary divisions where the t_2 and t_3 hubs are descendants of the t_1 hub; t_4 and t_5 - descendants of the t_2 hub; t_6 and t_7 - descendants of the t_3 hub.

3 CART algorithm on a set of data of training

Step 1. Choice of splitting.

Let's say that t is any subsets of a binary tree T minimum and square regressions; $\bar{d}(t)$ - this average d_i value for couples $d_i (x_i, d_i)$ which get to t sub tree:

$$\bar{d}(t) = \frac{1}{N(t)} \sum_{x_i \in t} d_i, \quad (1)$$

at the same time we make summation on all d_i for which $x_i \in t$, and $N(t)$ - total of similar cases.

Now we will define the following indicators:

$$E(t) = \frac{1}{N} \sum_{x_i \in t} (d_i - \bar{d}(t))^2 \quad (2)$$

$$E(t) = \sum_{t \in T} E(t) \quad (3)$$

For t the sum $\sum_{x_i \in t} (d_i - \bar{d}(t))^2$ is an intra nodal sum of squares or the general square deviation of all d_i of a sub tree of t from an average $\bar{d}(t)$.

Let's say what we have S splitting a hub t of a tree of T . The best splitting s^* will be that for which $E(t)$ accepts the smallest value, and at the same time will assume that for some splitting s of a hub t into t_R (knot to the right of t) and t_L (knot to the left of t) will be fairly following ratio:

$$\Delta E(s, t) = E(t) - E(t_L) - E(t_R). \quad (4)$$

The best splitting s^* will be that at which the constructed regression tree will maximize decrease of $E(t)$:

$$\Delta E(s^*, t) = \max_{s \in S} \Delta E(t, s). \quad (5)$$

Step 2. Identification of terminal knot.

The hub t is terminal if fairly following condition:

$$\max_{s \in S} \Delta E(t, s) < \beta, \quad (6)$$

where β - in advance defined threshold.

Step 3. Application of a method of the smallest squares for assessment of parameters of terminal knot.

Let's assume that t - terminal knot of a binary tree of T ; $X(t)$ - a matrix which is made from $x_i \in t$; $d(t)$ - the vector created by desirable responses for t sub tree. Then we will calculate a vector of $w(t)$:

$$w(t) = X^+(t)d(t), \quad (7)$$

where: $X^+(t)$ - matrix, which the pseudo-return to a matrix $X(t)$.

When using scales on a formula (7) the problem of search of splitting can be solved thanks to calculation of the smallest sum of squares of mistakes in relation to the surface of regressions (but not average values).

4 Initialization of the HME model by means of a CART algorithm

Some tree of decisions for an objective as a result of use of a CART algorithm in relation to a set of data of training is Let's say constructed. Splitting on a CART algorithm it is represent able as a multidimensional surface where x - an entrance vector; a - vector of parameters; b - the set threshold:

$$a^T x + b = 0. \quad (8)$$

The regression surface which is created by network of a lock in a binary tree can be described by the following formula (9) and represents splitting (at $g = 1/2$):

$$g = \frac{1}{1 + \exp(-(a^T x + b))}, \quad (9)$$

The a vector of scales for the network of a lock considered by us corresponds to equality:

$$a = \|a\| \cdot \frac{a}{\|a\|}, \quad (10)$$

where is $\|a\|$ the length of a vector a ; $a / \|a\|$ - the vector having single length. If to substitute (10) in (9), then we will receive parametrical splitting in network of a lock of the following look:

$$g = \frac{1}{1 + \exp\left(-\|a\| \left(\left(\frac{a}{\|a\|} \right)^T x + \frac{b}{\|a\|} \right)\right)}. \quad (11)$$

Follows from a formula that the vector $a/\|a\|$ defines the direction of splitting, and sets sharpness of splitting. Network of a lock in (11) which is created on the basis of a linear filter which *softmax* (nonlinear activation function) follows, is capable to imitate the splitting similar to CART splitting. Besides we receive length of vectors of parameter a - additional degree of freedom that exerts considerable impact on $\|a\|$ which is carried out by network of a lock in HME model. Respectively, the following properties are characteristic of a vector of synoptic scales a with the fixed orientation: with rather big length of a vector splitting will be sharp; with a small length of a vector splitting will be soft.

In case sharpness of splitting is equal to zero ($\|a\|=0$), then splitting will disappear also on both sides of mental splitting $g=1/2$. If length of a vector to accept equal to zero, then it is equivalent to removal of non-terminal knot from a tree (since the network of a lock does not contain splitting any more). If length of a vector is extremely small in all non-terminal knots, then the HME model functions as usual model of linear regression (like one knot). If length of a vector increases, then the HME model forms soft splitting, increasing thereby quantity of the degrees of freedom available to model.

Generalizing all aforesaid, initialization of HME model happens on the basis of the following algorithm:

Step 1. Application of a CART algorithm to the training data.

Step 2. Application of a vector of synoptic scales of experts of HME model to values of estimates (which are received by means of a method of the smallest squares for vectors of parameters of necessary terminal knots of use of a CART algorithm of a binary tree constructed later).

Step 3. A task of vectors of synoptic scales for networks of locks, orthogonally to the corresponding directions of splitting a binary tree which was received after use of a CART algorithm.

Step 4. Calculation of length of vectors of synoptic scales and their establishment by identical values for lengths of small casual vectors.

The architecture of model of hierarchical mixture of opinions of experts (HME model) is similar to architecture of model of a tree of classification and regressions (CART model), but favorably differs from the last in softness of splitting entrance space. A CART algorithm, being a simple example of model, applies tough decisions at a stage of splitting entrance space into a set of subspaces with own experts that inevitably leads to loss of some information.

The HME model is similar to model of a multilayered perception (MLP - multilayer perception) as it applies the enclosed nonlinearity form, but, unlike MLP model not for formation of display of an entrance signal in day off (that demands approach of "a black box" for creation of uniform approximating function for data of training and leads to loss of depth of a view of an initial task), and for splitting entrance space.

5 Conclusions

Thus, the HME model is an associative car with dynamic type which combines advantages of both CART model, and MLP.

This model allows to come to the common decision in complex computing challenges of creation of optimum information management systems city development combining the constructed compromise of opinions of a great number of experts due to some simplification and splitting entrance space by experts into a set of subspaces.

References

1. M.I. Jordan, The Eighth Annual Conference of the Cognitive Science Society, 531-546 (1986)
2. D.N. Osherson, S. Weinstein, M. Stoli, *Modular learning* (MIT Press, Cambridge, 1990)
3. V.E. Belousov, I.P. Abrosimov, O.V. Gubina, Bulletin of VSU. Series: System analysis and information technologies **3**, 124-129 (2017)
4. A.A. Galinskaya, Mathematical cars and systems **3(4)**, 87-102 (2003)
5. T. Kormen, Ch. Leyzerson, R. Rivest, K. Stein, *Algorithms: construction and the analysis* (Williams, 2005)