Two-Level Model Of Heterogeneous Ensembles Of Neural Networks For The Analysis Of Telemetry Data

Yauheni Marushko, United Institute of Informatics Problems of National Academy of Sciences of Belarus

Abstract
This article gives a brief description of the main methods of forming parallel ensembles of experts, in particular ensembles of neural networks. Also the learning algorithm of neural network ensembles with elements of the evolution strategy described. And the model of two-level training is given. This model is based on an ensemble of neural networks that use mentioned algorithm of learning. Also presents the results of tests of the model on the task of forecasting multivariate time series of telemetry data that describe the state of small airborne objects.

1. Introduction
Telemetry covers the measurement of physical quantities characterizing the state of objects or processes, the transfer of the results of these measurements, recording and data processing.

The processing and analysis telemetry data in a continuous process is accompanied by noisy data. This usually creates by non-deterministic sources of noise. This makes it the preferred use of the technology of artificial neural networks.

The following benefits are the basis of the development of systems to analyze telemetry data using a neural network [1]:

- The ability for self-learning, the replacement of complex mathematical apparatus necessary volume of information, the initial function of its transformation and the structure of the neural network;
- The possibility of realizing nonlinear mappings with substantial non-linearities in the hidden neurons of multilayer neural networks;
- High degree of parallelism possible for a neural network software system in real time.

These advantages provide application of neural network algorithms in the diagnosis of complex dynamic objects. Neural networks appear here as the apparatus of formalization of complex algorithms that convert the information.

There are examples of the use of neural networks in on-board intelligent decision support systems for managing complex dynamic objects and diagnosis of its condition [2].

The problem of forecasting requires an individual approach to processing data when using traditional neural networks. Such an approach requires a lot of time and resources, as determined during the design of a large number of parameters of both the neural network model and the processing parameters of time series (prediction window, the time horizon, etc.). It is therefore necessary to develop a tool automatic settings and automatic choice of architecture. Such a mechanism can be implemented using neural network complexes [3, 4, 5]. There are several examples of such complexes, they are mainly designed to solve a particular practical problem, and have restrictions on use.

So an important task is to develop a neural network structure, which allows you to automatically search for the optimal solution. Such structures can be used to determine the preferred structure of a neural network, or to determine the individual characteristics of the structure and process. Most often, these structures are implemented based on ensembles of neural networks, and therefore also open to question is the optimization of learning algorithms.

2. Ensembles of neural networks (ENN)
Algorithms for the separation of complex learning tasks into many separate sub-tasks are used to improve the accuracy and to use the simultaneous processing of large data sets. Further, individual solutions are combined in ensembles of models.

Two criteria must be optimized during the formation of the ensemble of models - the quality of training of particular network and optimal integration of models.

2.1 Methods for combining experts
There are several basic methods of combining independent models in ensembles: Bagging, Stacking, specialization of experts, maps of expert and mixture of experts [6,7, 8, 9].

2.2 Combining using bootstrap
Same model are varied at training on various data samples. If we have only one set of m examples, the
different sub-samples with a similar statistic can be obtained by applying the bootstrap - a random sample with return [7].

The approach is based on the independent learning of individual models on bootstrap samples from the training data set and then combining the obtained models in the committee majority (ensemble). The method is called "beggings" (bagging - bootstrap aggregating).

Begging method shows stable results in practice, and easy to implement. The spread of the output values of different experts may be used for further evaluation error (the upper bound). For the sequential computation committee may expand in stages, with the first results of the classification system begins to produce after the first training models.

2.3 Stacking

The algorithm uses as the base models different classification (prediction) algorithms that are trained on the same data. Then, a meta-classifier (supervisor) is trained on the input data, supplemented by basic algorithms. The supervisor can use when teaching the estimates of parameters of the distribution (for example, estimates the probability of each class) instead of the results of basic algorithms [8].

Thus, meta-classifier is trained to distinguish which of the basic algorithms to be "trusted" on certain input data. Stacking algorithm is trained using cross-validation.

2.4 Specialization of experts

The method assumes that after training all of the examples are stored along with the estimates of errors that were made to them by all the members of the ensemble. Then, after receiving a new example, it matches closest prototype of the training set in the database and in the vote involving only those members of the ensemble, which allow small errors found on prototypes. Thus, there is a dynamic union of some members of the committee for solving each new task [6].

2.5 Maps of expert

Dynamic aggregation can be extended if, instead of storing the prototypes use a cluster structure (e.g., Kohonen maps), which is built on the training set without tags classes and assign to each cluster a list of models that show the best results on the data of this cluster. For the classification is now sufficient to find the nearest cluster and apply its associated model [6].

Each member is assigned to all clusters in the data from which it has a small error, thus the clusters maintain their rating list of models. Neural map of specialized expertise can be used not only in the process of recognition, but for analytical purposes.

In particular, based on the map as requests, we can define the "problem areas" of the ensemble. These clusters are filled with data, but the poor are associated experts. This gives a new training strategy and replenishment of the ensemble with new models.

2.6 Mixtures of experts

Combining experts in the processing of a new sample is made with weights that are determined by the control outputs of the neural network. The task of the control neural network - choose the weight of the mixture, the most relevant being processed at the moment the input vector [9].

In various versions of the model as a control network developers use the radial basis function, the system of fuzzy rules, hierarchical structures and other models.

Mixtures of experts are an example of a nonlinear association of experts in the ensemble. In general, it is difficult - obtain the theoretical constructs that describe the accuracy of such models. In practice, the usual methods of estimation and cross-validation are used [6, 9].

3. Training ensembles of neural networks

Parallel training of ensembles can be optimized by using the elements of evolutionary approaches. The basic operation here is a mutation where in which the connection weight is perturbed by random value with a certain probability.

Also one of variants crossover can be applied if for every single model can identify the "genotype".

The algorithm of formation of population consists of the following steps [11]:

1. Initializing weights of each neural network from ensemble.
2. Mutation of weighting links, applied crossover if possible, training.
3. Survival of the fittest. Survival implies the elimination of a set of less adapted experts, and duplication of the fittest.
4. 5. Go to (2) or stop if stopping criteria met.

In 3) the population of networks is ranked in descending order by relative fitness. In 4) part of the fittest networks are replicated twice and inserted into the next population, the next part are replicated once and inserted. The bottom 20% is discarded from the population. This is then repeated until the criteria for stopping is met. Elitism is also implemented, such that one of the two replicas of the fittest individual Neural Network passed into the next generation does not have its weights perturbed.

4. Two-level learning model

Using only a single ensemble at designing of a forecasting model often does not give the desired
result, and leads to development that is repeated over and over again with varying architecture, varying the types and learning algorithms of neural networks [5]. This is due to the fact that for some data better use one type or architecture, for other data - the other type, and architecture. It is proposed to use a two-level learning model [11]. This model allows realizing heterogeneity of the complex neural network (see Fig. 1).

The first level of structure is a set of ensembles of heterogeneous networks. First, it can be different types of neural networks [3, 12, 13]. Second, it may be similar networks (may have a similar architecture and parameters), but with different time-series analysis [4]. So the elements of this model may use different delays between the window and the predicted value of forecasting, to determine the delay in data being processed. Also, time series can be processed in a different scale, so each element will be fed data at a different time scale. This architecture may account for distant effects.

Third, it may be similar networks with different training parameters. This architecture can be used to find the optimal parameters presented neural network model.

The second level is a hybrid ensemble which is formed of the most successful elements of the ensembles of the first level after passing a certain number epochs of training.

Also, as a second level of the ensemble can be used the ensemble (or a single network) supervisor [4, 7, 10], processing the output values of all elements of the first level.

Fitness the first-level ensembles are analyzed for forming hybrid ensemble. If one of the ensembles is much more suited than others, the hybrid ensemble is formed only on this basis. In other cases, a hybrid ensemble made up of the fittest neural networks most accurate ENNs.

![Fig. 1 - Two-level learning model.](image)

5. The experimental part

For forecasting, we used data from six aircraft sensors that measure the following parameters: distance, speed, heading angle on the first line of communication (FLC), pitching angle of FLC, heading angle on the second line of communication (SLC) and pitching angle according to SLC.

We also used a set of «The Santa Fe Time Series Competition Data», synthetic generated multivariate time series of states of objects.

Telemetric information about the object was considered in two ways (see Fig. 3). In the first case, the data from the sensors are analyzed directly, and then the resulting forecast determines predicted state. In this case, training and processing is done on the initial data.

In the second case the input data is classified into a number of states of the object, which is analyzed, so it significantly reduces the the volume of data being processed. Comparison of predicted and actual data is shown in Fig. 4.

![Fig. 3 - Two approaches states of the objects data processing.](image)

In this case, we have not performed considerable pre-processing of data. But to get the best quality of results is also necessary to analyze the input data. It is necessary to analyze the correlation between different variables, to allocate the independent inputs, to determine the precursors, if they exist, to remove noise. Thus we can achieve significant reductions in processing information.

Evaluation of the accuracy of the model is also an important component of testing. For this purpose, a series of experiments was performed. The different
single neural network, different ENN, and the proposed model with different sets of ENN were trained on the same data. The results are shown in Table 2: the average error on time series (Er. 1), and on classified data (Er. 2).

<table>
<thead>
<tr>
<th>Model</th>
<th>Er. 1, %</th>
<th>Er. 2, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single ANN</td>
<td>6,7-24</td>
<td>5,5-22</td>
</tr>
<tr>
<td>ENN</td>
<td>5,6-17,2</td>
<td>2,7-14</td>
</tr>
<tr>
<td>The proposed model</td>
<td>5,6-8,3</td>
<td>2,7-7,7</td>
</tr>
</tbody>
</table>

6. Conclusion

In this paper, a generalization of existing models of complexes based on the ensembles of neural networks has been submitted. We also proposed the use of two-level training algorithm of heterogeneous ensembles, using training with elements of an evolutionary strategy for processing telemetry data of the states of small airborne objects.

The software module was developed that implements the proposed method and a series of experiments performed to determine the characteristics of software modules and models. The results showed that two-level learning algorithm allows to build neural network model for the given ENN close to optimal in the specified range of parameters. There is no need for a set of experiments to get a good result.

Disadvantages include increased training time, by using a set of single neural networks.

Bibliography


Authors:

Yauheni Marushko
United Institute of Informatics Problems of National Academy of Sciences of Belarus
st. Surganova 6, 220013 Minsk

email: marushko@gmail.com