UDC 004.855 (045)
DOI: 10.18372/1990-5548.55.12772

[1]O. I. Chumachenko,
[2]A. O. Kuzmenko

ALGORITHM OF PRUNING OF HYBRID NEURAL NETWORKS ENSEMBLE

[1]National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute," Kyiv, Ukraine
[2]Aviation Computer-Integrated Complexes Department, Educational & Research Institute of Information and Diagnostic Systems, National Aviation University, Kyiv, Ukraine
E-mails: [1]chumachenko@tk.kpi.ua [2]kuzmenkooao@gmail.com

***Abstract*—*Despite the fact that the ensemble is usually more accurate than a single network, existing ensemble techniques tend to create unreasonably large ensembles that increase the use of memory and computation costs. The ensemble's pruning solves this problem. The article analyzes the compromise between accuracy and diversity and it is proved that classifiers, which are more accurate and make more predictions in the minority group, are more important for the construction of the subensemble. A metric that takes into account accuracy and diversity is proposed to evaluate the contribution of a separate classifier that will help to allocate the required number of networks with the best results.*

Index Terms—Ensemble pruning; bagging; accuracy; diversity.

I. INTRODUCTION

The desirable variety of models and a certain independence of their errors, which positively affects the accuracy during the union into the ensemble, can be achieved without the use of some models to change others. Models can be independently learnt, which is a fundamental resource when using parallel computing. Equivalent models become diverse when learning on different sample data. Having only one set of m examples, it is possible to get subsamples with close statistics using the bootstrap [1] – random sample with return. This is the basis for an approach called bagging. The approach is based on the independent learning of individual models on bootstrap samples from the training set of data and the union of the resulting models in the ensemble of majority.

The construction of classical ensembles is an active research area in the communities of machine learning and intelligent data analysis. Rather than relying on one classifier, the ensemble is a set of classifiers that decide collectively [2]. A necessary and sufficient condition for a classifier ensemble to be more precise than any of its individual members is that the classifiers are precise and diverse. Since the diversity of the ensemble decreases with increasing accuracy of the members, the key to the success of any method of the ensemble learning is a compromise between precision and diversity.

This article provides an algorithm for pruning the ensemble that allows you to select the percentage of neural networks according to available resources and the best compromise between accuracy and diversity. The algorithm is based on the analysis of work [3].

II. PROBLEM STATEMENT

Let $D=\{d_1,...,d_2\}$ be a set of N data points where $d_i=\{(x_i,y_i)|i\in[1,N]\}$ is a pair of input features and label that represents the ith data point, $C=\{c_1,...,c_m\}$ be a set of M classifiers where $c_j(x_j)$ gives the prediction of the ith classifier on the jth data point, $V=\{V^{(1)},...,V^{(N)}\,|\,V^{(j)}=$ $=\left[V_1^{(j)},...,V_L^{(j)},i\in[1,N]\right]$ be a set of vectors where $v_j^{(j)}$ is the number of predictions for the jth label of the ith data point of an ensemble combined with majority voting, and L is the number of output labels. There is a data sample, which is divided by learning and testing. It is necessary to select networks $C=\{c_1,...,c_m\}$ for the formation of the ensemble on the basis of the accuracy and diversity of the classifiers, considering that the networks were previously learnt on bootstrap-samples, formed from the training set of data.

Given two classifiers c_i and c_j, where $N^{(01)}$ denotes the number of data points incorrectly predicted by c_i but correctly predicted by c_j, and $N^{(10)}$ is the opposite of $N^{(01)}$, the diversity of c_i and c_j, denoted by $\mathrm{Div}_{i,j}$, is the ratio between the sum of the number of data points correctly predicted by one of the classifiers only and the total number of data points, as given in equation

$$\mathrm{Div}_{i,j}=\frac{N^{(01)}+N^{(10)}}{N}. \qquad (1)$$

The contribution of the variety of the classifier c_j to the ensemble, being denoted as ConDiv_j, is the sum of the difference between c_j and each other classifier in the ensemble.

$$\mathrm{ConDiv}_j = \frac{1}{N}\sum_{k=1}^{N}\left(M - v^{(j)}_{c_i(x_k)}\right), \quad (2)$$

where N is the number of data points; M is the total number of classifiers; $v^{(j)}_{c_j(x_k)}$ is the number of classifiers that agree with the classifier in predicting the result of the classification.

The prediction of a member of an ensemble at one data point can be divided into four subsets:

1) a subset in which the classifier predicts correctly and is in the minority group;

2) a subset in which the classifier predicts correctly and is in the majority group;

3) subset, in which the classifier incorrectly predicts and is in a minority group;

4) subset, in which the classifier is incorrectly predicts and is in the majority group.

Therefore, heuristic metrics are proposed to evaluate the contribution of each member of the ensemble, which takes into account both accuracy and diversity. Thus, the inclusion of a network in an ensemble is determined on the basis of an individual contribution.

III. Individual Contribution Calculation

The individual contribution of the classifier c_j is determined as follows:

$$IC_i = \sum_{j=1}^{N} IC_i^{(j)}, \quad (3)$$

where $IC_i^{(j)}$ is the contribution of the classifier c_j into jth data point d_j.

$IC_i^{(j)}$ is determined depending on the subset, which includes the classifier prediction.

1) When $c_j(x_j)$ is equal to y_j, which means that c_j makes the correct predictions at the point d_j, if $c_j(x_j)$ belongs to a minority group, then it is defined as

$$IC_i^{(j)} = 2v^{(j)}_{\max} - v^{(j)}_{c_i(x_j)}, \quad (4)$$

where $v^{(j)}_{\max}$ is the number of votes in d_j and $v^{(j)}_{c_i(x_j)}$ is the number of predictions $c_j(x_j)$, calculated earlier.

2) When $c_j(x_j)$ is equal to y_j and $c_j(x_j)$ belongs to the majority group, it is defined as:

$$IC_i^{(j)} = v^{(j)}_{\mathrm{sec}}, \quad (5)$$

where $v^{(j)}_{\mathrm{sec}}$ is the second largest number of votes on labels d_j.

3) When $c_j(x_j)$ is not equal to y_j (two negative cases), $IC_i^{(j)}$ is defined as:

$$IC_i^{(j)} = v^{(j)}_{\mathrm{correct}} - v^{(j)}_{c_i(x_j)} - v^{(j)}_{\max}, \quad (6)$$

where $v^{(j)}_{\mathrm{correct}}$ is the number of votes for the correct label d_j.

Then equation (3) can lead to the following form, combining the equations (4), (5) and (6):

$$\mathrm{IC}_i = \sum_{j=1}^{N}\left(a_{ij}\left(2v^{(j)}_{\max} - v^{(j)}_{c_i(x_j)}\right) + \theta_{ij}\left(v^{(j)}_{\mathrm{correct}} - v^{(j)}_{c_i(x_j)} - v^{(j)}_{\max}\right)\right). \quad (7)$$

Thus, for each data point, an individual contribution is calculated that takes into account both the accuracy and the diversity.

IV. General Algorithm

The general algorithm can be represented as follows.

1) There is a set of learning examples $(x_1, y_1),\ldots,(x_m, y_m)$ with labels $y \in \{1,\ldots,k\}$.

2) Get t bootstrap samples D_t.

3) Independently (in parallel) to train t classifiers h_t, each on their sample D_t.

4) Obtain the predictions of each classifier c_j for the jth data point d_j on the test sample and determine the individual contribution.

4.1. If the prediction $c_j(x_j)$ equals y_j, that is, c_j makes the correct predictions in d_j

4.1.1. If the prediction $c_j(x_j)$ belongs to a minority group, then the individual contribution is calculated by the equation (4)

4.1.2. If the prediction $c_j(x_j)$ belongs to the majority group, then the individual contribution is calculated by the equation (5)

4.2. If the prediction $c_j(x_j)$ is not equal to y_j, the individual contribution is calculated by the equation (5)

5) Determine the individual contribution of the classifier c_j by the equation (7), where, depending on item 4, the correspondent coefficient is put into one.

6) Add the pair (c_j, IC_j) to the OL list and sort in descending order.

7) Determine the parameter p, which is the desired percentage of classifiers C, which should be stored at the output of the subassemble. This parameter is determined based on existing resources such as memory and time consuming.

8) Knowing the desirable cost of resources and real, bring out the first p percent of the list as a pruned and ranked subensemble. But all of these approaches are based on selection of networks by accuracy or diversity, which have already been considered during the ranking by individual contribution. It is also known that sometimes it is not enough to take into account diversity and accuracy in order to form an effective ensemble. It is proposed to use such an approach as Complementarity Measure, which also takes into account the interaction of classifiers among themselves.

V. JUSTIFYING THE NEED FOR PRUNING

As a result of the algorithm, we received the ranking of selected neural networks by their individual contribution. In practice, in assemblies, the classification error usually shows a monotonous decrease, as a function of the number of elements in the ensemble. For large size of the ensemble, these curves reach the asymptotic constant error level, which is usually considered the best result that can be achieved. Typical methods for selecting the required classifiers are Reduce-Error Pruning, Kappa Pruning, Margin Distance Minimization, and Orientation Ordering.

VI. PRUNING ALGORITHM

The pruning algorithm can be represented as follows.

1) Get predictions for each member of the ensemble after run on the test data set.

2) To form a subensemble of S_{u-1} from the members, which at the output have a prediction $c_j(x_j)$ that does not equal y_j (that is, they predict incorrectly).

3) Select a classifier with the best diversity and accuracy according to the resulting list OL as a result of the previous algorithm and add it to the subensemble.

4) Get the value S_u, which characterizes the impact of the classifier with the best results on the subensemble, which gives the wrong predictions:

$$S_u = \arg\max_k \sum_{(x,y)\in Z_{sel}} I(y = h_k(x) \text{ and } H_{s_{u-1}}(x) \neq y), \quad (8)$$

where the classifier with the best results belongs to the original ensemble $h_k \in E_t / S_{u-1}, I(\bullet)$ – indicator function (I(true) = 1 and I(false) = 0), S_u is the measure according to which the members are selected for the pruned subensemble.

5) It is necessary to set the threshold for the selection of the classifier. That is, if the error made by the subensemble of S_{u-1} is more than the S_u error, namely, the difference between their errors exceeds the predefined threshold, then the classifier is considered to be an addition and is taken into a pruned subensemble. So,

$$S_u = \arg\max_k \sum_{(x,y)\in Z_{sel}} I\big(|y - H_{s_{u-1}}(x)| - |y - h_k(x)| > \text{threshold}\big). \quad (9)$$

6) Repeat steps 3–5 with each classifier and the initial subensemble S_{u-1}.

IV. CONCLUSIONS

So, one of the problems of ensemble approaches is that they try to create unreasonably large ensembles, which requires a large amount of memory to store trained classifiers and reduce the response time for prediction. The pruning of an ensemble is a method that solves this problem by choosing a subset of individual classifiers from a prepared ensemble to form nodes for prediction. It is necessary to carefully select the classifiers in the subensemble so that it is small enough to reduce the memory requirements and the response time with the precision of the prediction, similar or with accuracy that exceeds the output ensemble.

REFERENCES

[1] Zhenyu Lu, Xindong Wu, Xingquan Zhu, and Josh Bongard, Ensemble Pruning via Individual Contribution Ordering. Burlington. 2007, 10 p.

[2] Gonzalo Martı´nez-Munoz, Daniel Herna´ndez-Lobato, and Alberto Sua´rez, "An Analysis of Ensemble Pruning Techniques Based on Ordered Aggregation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 2, 2009, pp. 245–259.

[3] Z.-H. Zhou, J. Wu, and W. Tang, "Ensembling Neural Networks: Many Could Be Better than All," *Artificial Intelligence*, vol. 137, pp. 239–263, 2002.

Received November 12, 2017

Chumachenko Olena. Candidate of Science (Engineering). Assosiate Professor.
Technical Cybernetic Department, National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute," Kyiv, Ukraine.
Education: Georgian Politechnic Institute, Tbilisi, Georgia, (1980).
Research area: system analysis, artificial neuron networks.
Publications: more than 60 papers.
E-mail: chumachenko@tk.kpi.ua

Kuzmenko Anastasia. 6th year of study student.
Aviation Computer-Integrated Complexes Department, National Aviation University, Kyiv, Ukraine.
Education: National Aviation University, Kyiv, Ukraine, (2015).
Research interests: automation of technological processes, neural networks.
Publications: 6.
E-mail: kuzmenkooao@gmail.com

О. І. Чумаченко, А. О. Кузьменко. Алгоритм спрощення гібридних нейронних мереж

Показано що ансамбль, зазвичай, більш точний, ніж одинична мережа, існуючі ансамблеві методи, однак, як правило, створюють невиправдано великі ансамблі, які збільшують використання пам'яті та обчислювальні витрати. Спрощення ансамблю вирішує цю проблему. Проаналізовано компроміс між точністю і різноманітністю і доведено, що класифікатори, які є більш точними і роблять більш точні прогнозів у групі меншості, більш важливі для побудови підансамблю. Запропоновано метрику, яка враховує точність і різноманітність, щоб оцінити вклад окремого класифікатора в ансамбль. Це дозволяє виділити необхідну кількість мереж з найкращими результатами і включити їх в ансамбль.
Ключові слова: спрощення ансамблю; беггінг; точність; різноманітність.

Чумаченко Олена Іллівна. Кандидат технічних наук. Доцент.
Кафедра технічної кібернетики, Національний технічний університет України «Київський політехнічний інститут ім. Ігоря Сікорського», Київ, Україна.
Освіта: Грузинський політехнічний інститут, Тбілісі, Грузія, (1980).
Напрям наукової діяльності: системний аналіз, штучні нейронні мережі.
Кількість публікацій: більше 60 наукових робіт.
E-mail: chumachenko@tk.kpi.ua

Кузьменко Анастасія Олександрівна. Студентка 6-го курсу.
Кафедра авіаційних комп'ютерно-інтегрованих комплексів, Національний авіаційний університет, Київ, Україна.
Освіта: Національний авіаційний університет, Київ, Україна, (2015).
Напрям наукової діяльності: автоматизація технологічних процесів, нейронні мережі.
Кількість публікацій: 6.
E-mail: kuzmenkooao@gmail.com

Е. И. Чумаченко, А. А. Кузьменко. Алгоритм упрощения гибридных нейронных сетей

Показано что ансамбль, как правило, более точный, чем единичная сеть, существующие ансамблевые методы, зачастую создают неоправданно большие ансамбли, которые увеличивают использование памяти и вычислительные затраты. Упрощение ансамбля решает эту проблему. Проанализирован компромисс между точностью и разнообразием и доказано, что классификаторы, которые являются более точными и делают более точные прогнозов в группе меньшинства, более важны для построения подансамбля. Предложена метрика, которая учитывает точность и разнообразие, чтобы оценить вклад отдельного классификатора в ансамбль. Это позволяет выделить необходимое количество сетей с лучшими результатами и включить их в ансамбль.
Ключевые слова: упрощение ансамбля; беггинг; точность; разнообразие.

Чумаченко Елена Ильинична. Кандидат технических наук. Доцент.
Кафедра технической кибернетики, Национальный технический университет Украины «Киевский политехнический институт им. Игоря Сикорского», Киев, Украина.
Образование: Грузинский политехнический институт, Тбилиси, Грузия, (1980).
Направление научной деятельности: системный анализ, искусственные нейронные сети.
Количество публикаций: более 60 научных работ.
E-mail: chumachenko@tk.kpi.ua

Кузьменко Анастасия Александровна. Студентка 6-го курса.
Кафедра авиационных компьютерно-интегрированных комплексов, Национальный авиационный университет, Киев, Украина.
Образование: Национальный авиационный университет, Киев, Украина, (2015).
Направление научной деятельности: автоматизация технологических процессов, нейронные сети.
Количество публикаций: 6.
E-mail: kuzmenkooao@gmail.com