

## Глава 4

### Лекция 11. Negative Correlation Learning

# Оглавление

<b>4 Лекция 11. Negative Correlation Learning</b>	<b>1</b>
11.1 Negative Correlation Learning . . . . .	2
11.1.1 Bagging . . . . .	2
11.1.2 Boosting . . . . .	2
11.1.3 Идея Negative Correlation Learning . . . . .	3
11.1.4 Эволюционный вариант Negative Correlation Learning . . . . .	4
11.2 Сравнение градиентного и эволюционного вариантов Negative Correlation Learning . . . . .	5

## 11.1 Negative Correlation Learning

Задачи классификации и аппроксимации можно решать не только с использованием 1 ИНС, но и с применением множества нейронных сетей, когда решение о значении аппроксимируемой переменной/классе принимается «коллективно». В этом случае говорят о комитете, или ансамбле ИНС, либо аппроксиматоров/классификаторов для более общего случая. Методы обучения множества классификаторов, работающих совместно, часто называют *комитетными*.

Метод, связанный с обучением ансамбля ИНС таким образом, чтобы их выходные сигналы как можно меньше коррелировали Negative Correlation Learning (NCL) разработан Зином Яо и Йонгом Лю [4]. Сам по себе этот подход не является эволюционным, ИНС обучаются традиционным градиентным алгоритмом, однако есть и разновидность, использующая эволюционное обучение [5].

Прежде чем приступить к описанию метода, рассмотрим ряд основных подходов к обучению комитетов классификаторов.

### 11.1.1 Bagging

Алгоритм Bagging (**B**ootstrap **a**ggregating) предложен Лео Бриманом [2]. По сути представляет собой простейший вариант комитетной классификации, когда решение принимается усреднением решений отдельных классификаторов. Можно показать, что в пределе, при очень большом числе классификаторов ошибка распознавания уменьшается [1].

Сами классификаторы обучаются с применением бутстрепинга (bootstrapping).

### 11.1.2 Boosting

Представляет более «продвинутой» метод, в котором каждый последующий классификатор обучается с учетом результатов обучения предыдущих классификаторов

[6]. В базовом варианте обучается 3 классификатора: первый  $C_1$  на случайном подмножестве данных, второй  $C_2$  на данных, из которых  $C_1$  правильно распознает лишь половину объектов, а  $C_3$  обучается на примерах, на которых ошибаются и  $C_1$  и  $C_2$ . Решение принимается по результатам голосования по принципу большинства: если  $C_1$  и  $C_2$  противоречат друг другу, то «на помощь» приходит  $C_3$ . Этот вариант работает для 2 классов.

Более продвинутый вариант алгоритма, AdaBoost [3], рассматривает обучение классификаторов с приоритетом для данных, на которых ошибались предыдущие классификаторы, причем общее число классификаторов может быть больше 3. При этом итоговое решение выносится взвешенным голосованием, веса классификаторов в котором определяются по результатам обучения.

### 11.1.3 Идея Negative Correlation Learning

Рассматривается ансамбль ИНС. Выходной сигнал вычисляется по формуле:

$$F(n) = \frac{1}{M} \sum_{i=1}^M F_i(n)$$

где  $F_i(n)$  – выход  $i$ -й ИНС,  $M$  – количество ИНС.

В целевую функцию вводится дополнительный штраф за коррелированность выходных сигналов ИНС:

$$E_i = \frac{1}{N} \sum_{n=1}^N E_i(n) = \frac{1}{N} \sum_{n=1}^N \frac{1}{2} (F_i(n) - d(n))^2 + \frac{1}{N} \sum_{n=1}^N \lambda p_i(n) \rightarrow \min, \quad (11.1)$$

где  $E_i(n)$  – ошибка  $i$ -й ИНС на  $n$ -м обучающем примере,  $d(n)$  – требуемый выходной сигнал для  $n$ -го обучающего примера,  $N$  – размер обучающей выборки,  $\lambda$  – параметр, регулирующий влияние штрафа.

Второе слагаемое в (11.1) вычисляется следующим образом:

$$p_i(n) = (F_i(n) - F(n)) \sum_{j \neq i}^N F_j(n) - F(n). \quad (11.2)$$

Минимизация ЦФ для ансамбля соответствует минимизации ЦФ для каждой сети в отдельности.

В базовом варианте алгоритма NCL обучение ИНС производится традиционным градиентным способом [4]. Кратко суть сводится к следующему: при обучении необходимо вычислить ошибку каждой ИНС и скорректировать ее выходной сигнал. Величина коррекции пропорциональна

$$-\frac{\delta E_i(n)}{\delta F_i(n)}.$$

Можно показать, что:

$$\frac{\delta E_i(n)}{\delta F_i(n)} = (1 - \lambda)(F_i(n) - d(n)) + \lambda(F(n) - d(n)). \quad (11.3)$$

Анализ выражений (11.1-11.3) позволяет сделать следующие выводы [4]:

1. В процессе обучения каждой сети сокращается не только  $F_i(n) - d(n)$ , но и  $F(n) - d(n)$ , таким образом учитываются ошибки других ИНС.
2. При  $\lambda = 0$ , штраф отсутствует и каждая сеть обучается независимо, т.е. индивидуальное обучение ИНС можно рассматривать как частный случай NCL.
3. При  $\lambda = 1$  из (11.3) получим

$$\frac{\delta E_i(n)}{\delta F_i(n)} = F(n) - d(n).$$

В этом случае можно показать, что

$$\frac{\delta E_i(n)}{\delta F_i(n)} \propto \frac{\delta E_{ens}(n)}{\delta F_i(n)}.$$

где  $E_{ens}$  – среднеквадратичная ошибка всего ансамбля ИНС на  $n$ -м обучающем примере:

$$E_{ens}(n) = \frac{1}{2} \left( \frac{1}{M} \sum_{i=1}^M F_i(n) - d(n) \right)^2.$$

Другими словами при  $\lambda = 1$  обучение отдельных ИНС ведет к уменьшению ошибки всего ансамбля, и NCL можно рассматривать как вариант декомпозиции исходной задачи обучения ИНС на  $M$  подзадач.

#### 11.1.4 Эволюционный вариант Negative Correlation Learning

Несмотря на то, что алгоритм NCL успешно используется и при градиентном обучении, существует его «эволюционный» вариант:

1. Инициализация популяции из  $M$  ИНС.
2. Обучение каждой ИНС в течение фиксированного количества поколений
3. Случайный выбор  $n_b$  родительских ИНС для порождения  $n_b$  потомков
4. Добавление потомков в популяцию и обучение. Веса других ИНС – "заморожены".
5. Вычисление приспособленности ИНС и селекция  $M$  лучших
6. Если алгоритм не закончен, то перейти на Шаг 3.
7. Формирование видов с помощью алгоритма  $k$ -средних.
8. Комбинирование видов для формирования ансамблей.

### Вычисление приспособленности

Используется разделение приспособленности.

Расстояние между ИНС определяется по примерам, которые эти ИНС правильно классифицируют.

Приспособленность определяется как сумма значений  $f_i$ .  $f_i$  равняется  $1/p_i$  или 0. Здесь  $p_i$  – количество особей, которые правильно классифицируют  $i$ -й обучающий пример.  $f_i = p_i$  если текущая особь правильно распознает  $i$ -й пример. Если особь распознает пример неправильно или таких особей вообще нет, то к приспособленности прибавляется 0.

### Мутация

Все особи мутируют. Вес связи меняется на величину, распределенную как  $\mathcal{N}(0, 1)$ .

### Формирование ансамбля

Ансамбль ИНС формируется из особей популяции в последнем поколении. Размер ансамбля желательно оставить небольшим. Два способа:

1. Все особи из популяции.
2. По одному представителю каждого вида в популяции.

Виды определяются с помощью алгоритма  $k$ -средних, в котором кластеризация ведется по выходным сигналам ИНС.

Комбинирование ИНС внутри ансамбля:

- Усреднение.
- Голосование.
- WTA.

## 11.2 Сравнение градиентного и эволюционного вариантов Negative Correlation Learning

Для сравнения использованы результаты из [4] и [5].

Алгоритм	Точность классификации
NCL, усреднение, $\lambda = 0$	13.86
NCL, усреднение, $\lambda = 1$	13.37
NCL, WTA, $\lambda = 0$	13.84
NCL, WTA, $\lambda = 1$	11.95
EENCL, популяция, усреднение	14.50
EENCL, популяция, WTA	13.50
EENCL, ансамбль, WTA	13.20

Таблица 4.1: Сравнение точности классификации на тестовом множестве данных для задачи «Australian credit card».

# Литература

- [1] C. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [2] L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.
- [3] Y. Freund and R. E. Schapire. Decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):pp. 119–139, 1997.
- [4] Y. Liu and X. Yao. Ensemble learning via negative correlation. *Neural Networks*, 12(10):1399–1404, December 1999.
- [5] Y. Liu, X. Yao, and T. Higuchi. Evolutionary ensembles with negative correlation learning. *IEEE Transactions on Evolutionary Computation*, 4(4):380–387, November 2000.
- [6] R. E. Schapire. The strength of weak learnability. *Machine Learning*, 5(2):pp. 197–227, 1990.