05.13.18

[1]Ф.А. Галимянов, [2]П.В. Малов

[1]Казанский государственный энергетический университет,
кафедра Информатики и информационно управляющих систем,
[2]Казанский национальный исследовательский технологический университет,
кафедра Информатики и прикладной математики,
Казань, fanisgalimyanov@gmail.com, pavel.malov@mail.ru

ВЕРОЯТНОСТНО-КОРРЕКТНАЯ В СМЫСЛЕ АППРОКСИМАЦИИ МОДЕЛЬ ОБУЧЕНИЯ

В статье рассматривается принцип обучения нейронной сети в системах двоичной классификации. В работе вводиться специальные термины, такие как: среда, понятия, класс понятий. Под понятием здесь следует понимать любой объект из предметной области. Используются примеры, которые могут быть как положительным, так и отрицательным, в зависимости от того входят они в понятия или нет. Приводится зависимости последовательности данных обучающей длины для понятия. Находиться вероятности ошибки при обучении с использованием целевой функции. Для алгоритма обучения, определяется два параметра управления: параметр ошибки и параметр доверия. В работе также приводится блочная диаграмма модели обучения.

Ключевые слова: *аппроксимация, вероятность, нейронные сети, двоичная классификация, обучение с учителем, класс понятий.*

Введение

Вероятностно-корректная в смысле аппроксимации (probably-approximately correct) модель обучения (PAC) была описана в [1]. Как и следует из ее названия, эта модель представляет собой вероятностный "каркас" (или среду) для изучения процессов обучения и обобщения в системах двоичной классификации. Она тесно связана с принципом обучения с учителем.

Постановка задачи

Сначала определимся с терминологией, связанной со средой X. Множество из элементов X называется понятием (concept), а любой набор его подмножеств — классом понятий (concept class). Примером понятия называется любой объект из предметной области, вместе с меткой своего класса. Если пример относится к данному понятию, он называется положительным примером (positive example). Если он не относится к данному понятию, то называется отрицательным примером (negative example). Понятие, для которого приводятся примеры, называется целевым (target concept). Последовательность данных обучения длины N для понятия ***c*** можно определить следующим образом:

$$T = \{(x_i, c(x_i))\}_{i=1}^{N} \tag{1}$$

В этой последовательности могут содержаться и повторяющиеся примеры. Примеры $x_1, x_2, \dots, x_n$ выбираются из среды X случайным образом, в соответствии с некоторым фиксированным, но неизвестным распределением вероятности. В определении (1) заслуживают внимания также следующие вопросы.

- Целевое понятие $c(x_i)$ рассматривается как функция, отображающая X в множество $\{0, 1\}$. При этом предполагается, что функция $c(x_i)$ неизвестна.
- Предполагается, что примеры статистически независимы. Это значит, что функция плотности совместной вероятности двух различных примеров x_i и x_j равна произведению соответствующих функций плотности вероятности.

В контексте терминологии, среда X соответствует пространству входных сигналов нейронной сети, а целевое понятие — ожидаемому отклику сети.

Набор понятий, порождаемых средой X, называется пространством понятий B. Например, пространство понятий может содержать фразы типа «буква А», «буква Б» и т.д. Каждое из этих понятий может быть закодировано различными способами при формировании множеств положительных и отрицательных примеров. В ракурсе обучения с учителем используется другое множество понятий. Обучаемая машина обычно представляет собой множество функций, каждая из которых соответствует определенному состоянию. Например, машина может предназначаться для распознавания «буквы А», «буквы Б» и т.д. Множество всех функций (т.е. понятий), определяемых обучаемой машиной, называется пространством гипотез (hypothesis space) G. Это пространство может совпадать или не совпадать с пространством понятий B. С определенной точки зрения пространства понятий и гипотез являются аналогами функции $f(x)$ и аппроксимирующей функции $F(x, w)$.

Предположим, что существует некоторое целевое понятие $c(x) \in B$, принимающее значения 0 и 1. Требуется обучить этому понятию нейронную сеть при помощи ее настройки на множестве данных T, определенном выражением (1). Пусть $\boldsymbol{g}(\boldsymbol{x}) \in \boldsymbol{G}$ — гипотеза, соответствующая отображению входа на выход, сформированному в результате проведенного обучения. Одним из способов достижения успеха в обучении является измерение степени близости гипотезы $g(x)$ к целевой концепции $c(x)$. Естественно, всегда существуют ошибки, обеспечивающие различие этих величин. Эти ошибки являются следствием того, что мы пытаемся обучить нейронную сеть некоторой функции на основе ограниченной информации о ней. Вероятность ошибки обучения определяется выражением

$$v_{train} = P\big(x \in X \colon g(x) \neq c(x)\big) \qquad (2)$$

Распределение вероятности в этом примере должно быть таким же, как и при формировании примеров. Целью обучения PAC является минимизация значения v_{train}. Предметная область, доступная алгоритму обучения, определяется размером N обучающего множества T. Кроме того, алгоритм обучения имеет два следующих параметра управления.

- Параметр ошибки (error parameter) $\varepsilon \in (0,1]$. Этот параметр задает величину ошибки, при которой аппроксимация целевого понятия $c(x)$ гипотезой $g(x)$ считается удовлетворительной.
- Параметр доверия (confidence parameter) $\delta \in (0,1]$. Этот параметр задает степень правдоподобия при построении "хорошей" аппроксимации.

Модель обучения PAC изображена на рисунке 1:

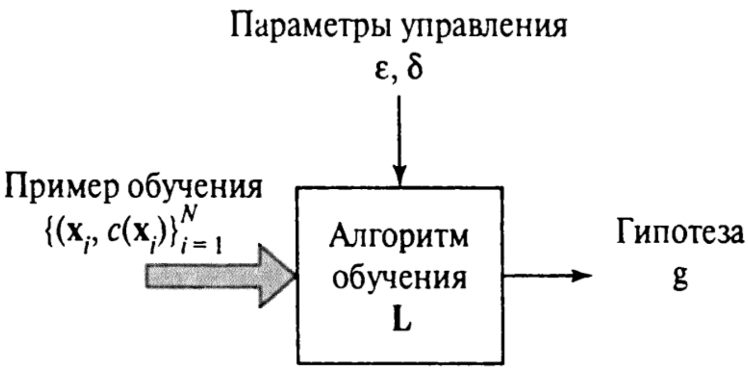

Рис. 1 - Блочная диаграмма модели обучения PAC

Выводы

Теперь можно формализовать модель обучения PAC [2], [3], [4].

Пусть B — класс понятий для среды X. Считается, что класс B является PAC - обучаемым, если существует алгоритм L, обладающий следующим свойством. Для любого целевого понятия $\boldsymbol{c} \in \boldsymbol{B}$, для любого распределения вероятности на X и для всех $\mathbf{0} < \boldsymbol{\varepsilon} < {}^{\mathbf{1}}/_{\mathbf{2}}$ и $\mathbf{0} < \boldsymbol{\delta} < {}^{\mathbf{1}}/_{\mathbf{2}}$ при использовании алгоритма L для множества примеров обучения $\boldsymbol{T} = \{(\boldsymbol{x_i}, \boldsymbol{c}(\boldsymbol{x_i}))\}_{\boldsymbol{i=1}}^{\boldsymbol{N}}$ вероятностью не хуже $(\mathbf{1} - \boldsymbol{\delta})$ результатом алгоритма обучения L будет гипотеза g с ошибкой обучения $\boldsymbol{v_{train}} < \boldsymbol{\varepsilon}$. Эта вероятность получается на любом случайном подмножестве множества T и при любой внутренней рандомизации, которая может существовать в алгоритме обучения L. При этом размер обучающего множества N должен превышать значение некоторой функции от $\boldsymbol{\delta}$ и $\boldsymbol{\varepsilon}$.

Другими словами, если размер N обучающего множества T достаточно велик, то существует вероятность, что в результате обучения сети на этом наборе примеров отображение входа на выход, реализуемое сетью, будет "приблизительно корректным". Обратим внимание, что, несмотря на зависимость от $\boldsymbol{\delta}$ и $\boldsymbol{\varepsilon}$, количество примеров N не обязательно зависит от целевого понятия с и распределения вероятности в X.

Список литературы

1. *Valiant L.G.* A theory of the learnable // Communications of the Association for Computing Machinery, 1984, vol. 27, p. 1134-1142

2. *Kearns M.J. and U.V. Vazirani* An Introduction to Computational Learning Theory // Cambridge, MA: MIT Press, 1994.

3. *Галимянов Ф.А., Малов П.В.* Конструктивные, независимые от распределения пределы обобщающей способности обучаемых машин // Научно-технический вестник Поволжья, № 6, 2020, с. 102-105.

4. *Vidyasagar M.* A Theory of Learning and Generalization // London: Springer-Verlag, 1997