

К ВОПРОСУ ПОСТРОЕНИЯ ЭФФЕКТИВНЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ДАННЫХ, ОПИСЫВАЕМЫХ РАЗНОТИПНЫМИ ПРИЗНАКАМИ

Н. А. ИГНАТЬЕВ

Национальный университет Узбекистана, Ташкент

e-mail: alimov@patent.uz

Some ways for synthesis of neural networks with the minimum configuration for solving recognition problem on the data with varied attributes are suggested.

Введение

Трудность реализации нейронных сетей (НС) в разнотипном признаковом пространстве в первую очередь связана с выбором весов входных параметров, измеряемых в слабых шкалах. Унификация данных путем перехода из сильных шкал измерения к слабым уменьшает количество используемой информации, делает выборки объектов комбинаторно ограниченными.

Наличие определенной априорной информации и свойств самих обрабатываемых данных дает возможность проводить так называемое “обогащение” слабых шкал, т. е. придавать количественные свойства шкалам порядка и наименования [1].

Процесс обучения НС в большинстве известных случаев проводится для заранее заданного фиксированного числа нейронов путем последовательного уточнения весов взвешенных сумм каждого из них. Процесс синтеза НС с минимальным числом нейронов теоретически плохо обоснован, его описания в литературе весьма поверхностны. Минимально необходимое число нейронов определяется с использованием процедуры контрастирования [2, 3], в основу которой положена оценка значимости весов связей в НС.

В [4] отмечалось, что для каждой прикладной области существует свой набор критериев, которые следует применять при выборе структуры НС. Получены формулы расчета степени обучаемости (пластичности) многослойных сетей прямого распространения, согласно которым обучаемость многослойных сетей не может быть выше, чем обучаемость однослойного персептрона.

Рассматриваемый в данной статье процесс синтеза НС с минимальным числом нейронов напрямую зависит от структуры признакового пространства описаний допустимых объектов. Отбор оптимального количества нейронов сводится к решению задачи минимального покрытия обучающей выборки объектами-эталоном.

1. Постановка задачи и метод решения

Считается, что множество допустимых объектов $E_0 = \{S_1, \dots, S_m\}$ представляет обучающую выборку, в которой заданы представители l непересекающихся классов K_1, \dots, K_l . Каждый допустимый объект описывается с помощью n признаков, r из которых количественные ($0 \leq r \leq n$), $(n - r)$ — номинальные.

Математическая интерпретация выбора весов признаков для нейронов сети выражается в построении кусочно-линейных дискриминантных функций, на базе которых принимается решение о принадлежности объектов к классам. Каждая дискриминантная функция отвечает за принятие решения по “своей” части признакового пространства.

Для определения по НС принадлежности произвольного допустимого объекта S к классам K_1, \dots, K_l необходимо:

а) вычислить множество весов входных параметров нейронов;

б) выделить минимальное покрытие множества E_0 объектами-эталонами, с помощью которых следует реализовать алгоритм, корректно распознающий объекты обучения.

Пусть I, J — множества номеров соответственно количественных и номинальных признаков, используемых для описания допустимых объектов, и объект $S_j \in E_0$ ($S_j = (x_{j1}, \dots, x_{jn})$) является эталоном выборки. Значение весов количественных признаков эталона вычислим аналогично [5, с. 92], т. е. $\forall t \in I, w_{jt} = x_{jt}$ и $w_{j0} = -\frac{1}{2} \sum_{t \in I} w_{jt}^2$.

Вычисление значений весов номинальных признаков по одному из предлагаемых способов основывается на предположении об одинаковом характере различий между количественными и номинальными признаками, т. е. считается, что максимальное различие объектов по количественным признакам соответствует максимальному возможному различию по каждому номинальному признаку. С помощью значений

$$w_{\max} = \max_{S_j \in E_0} (-2w_{j0}/r),$$

$$\lambda_{\max} = \sum_{t=1}^l |K_t| (|K_t| - 1),$$

$$\beta_{\max} = \sum_{t=1}^l |K_t| (m - |K_t|)$$

вес каждого номинального признака $c \in J$ определяется как

$$w_{jc} = \left(\frac{\lambda_c}{\lambda_{\max}} \right) \left(\frac{\beta_c}{\beta_{\max}} \right) w_{\max}, \quad (1)$$

где

$$\lambda_c = \sum_{j=1}^l \sum_{S_i, S_d \in K_j} \begin{cases} 1, & x_{ic} = x_{dc}, \\ 0, & x_{ic} \neq x_{dc}; \end{cases}$$

$$\beta_c = \sum_{j=1}^l \sum_{S_i, S_d \in E_0 \setminus K_j} \begin{cases} 1, & x_{ic} \neq x_{dc}, \\ 0, & x_{ic} = x_{dc}. \end{cases}$$

Значение взвешенной суммы по объекту-эталону $S_j \in E_0$ для произвольного допустимого объекта $S = (a_1, \dots, a_n)$ вычисляется как

$$\varphi(S, S_j) = \sum_{i \in I} w_{ji} a_i + \sum_{i \in J, x_{ji} = a_i} w_{ji} + w_{j0}. \quad (2)$$

Согласно решающему правилу, реализующему принцип “победитель забирает все”, объект S принадлежит к тому классу, значение взвешенной суммы (2) объекта-эталона которого максимально.

В связи с разномасштабностью измерений количественных признаков в определенных случаях целесообразно проводить их нормирование. Одной из форм нормирования является отображение значений признаков в $[0,1]$ по формуле

$$x_{ji}^* = \frac{x_{ji} - x_{\min}^i}{x_{\max}^i - x_{\min}^i}, \quad (3)$$

где x_{\max}^i, x_{\min}^i — соответственно максимальное и минимальное значения признака $i \in I$ на E_0 . Отказ от предположения об одинаковом характере различий количественных и номинальных признаков, а также нормирование по (3) делают ненужным вычисление значения w_{\max} и позволяют взамен (1) использовать формулу

$$w_{jc} = \left(\frac{\lambda_c}{\lambda_{\max}} \right) \left(\frac{\beta_c}{\beta_{\max}} \right). \quad (4)$$

Формулой (4) можно пользоваться и в тех случаях, когда описание допустимых объектов определяется только значениями номинальных признаков.

Поиск минимального покрытия F обучающей выборки эталонами с целью выбора минимальной конфигурации нейронной сети осуществляется процедурой “последовательное исключение”. Изначально все объекты E_0 считаются эталонами покрытия F , т. е. $F = E_0$. Если при использовании $F \setminus S_i$ в качестве эталонов алгоритм распознавания является корректным (не делающим ошибок) на E_0 , то производится удаление объекта S_i из F . Если исключение любого объекта-эталона из F и последующее использование F -алгоритма распознавания приводят к появлению ошибок на E_0 , то выполнение процедуры останавливается.

Итогом работы процедуры “последовательное исключение” является множество F , содержащее минимальное число объектов-эталонов, достаточное для корректного разделения объектов обучения на классы. Очевидно, что состав множества F зависит от порядка подачи объектов-кандидатов на удаление из числа эталонов.

Другой формой работы с разнотипными признаками является перевод номинальных признаков к булевым и последующее использование математических методов анализа количественных признаков [6] в объединенном пространстве количественных и булевых признаков. Обозначим через T множество номеров булевских признаков, мощность которого равна сумме градаций каждого номинального признака. Если признак $c \in J$ имеет q возможных значений $\{b_c^1, \dots, b_c^q\}$, то каждому b_c^i сопоставляется булев признак z_t с $t \in T$ как

$$z_t = \begin{cases} 1, & \text{если } x_c = b_c^i, \\ 0, & \text{если } x_c \neq b_c^i. \end{cases}$$

Далее в объединенном пространстве количественных и булевых признаков для минимального покрытия обучающей выборки эталонами можно использовать конечно-сходящийся алгоритм [7] или процедуру “последовательное исключение”.

В целях сравнительного анализа перечислим рассмотренные способы задания структуры данных для выбора на них минимальной конфигурации НС:

- 1) вычисление весов номинальных признаков по (1) с сохранением исходных значений количественных признаков;
- 2) преобразование количественных признаков в $[0,1]$, и вычисление весов номинальных признаков по (4);
- 3) перевод номинальных признаков в булевы и дальнейшее использование их в объединенном пространстве как количественных.

В первом способе уязвимым местом является использование предположения о зависимости весов количественных и номинальных признаков. Второй способ избавлен от предположений о зависимости весов количественных и номинальных признаков, зато требует предварительных вычислений максимальных и минимальных значений по каждому количественному признаку. К числу существенных недостатков третьего способа можно отнести рост размерности признакового пространства описаний допустимых объектов, что может привести к значительному увеличению затрат вычислительных ресурсов при синтезе НС.

Для численной оценки разных способов выбора структуры нейронных сетей, описываемых в данной работе, предлагается использовать меру статистического разнообразия класса кусочно-линейных решающих функций [8]. Предпочтение нужно отдавать той структуре нейронной сети, для которой

$$k * n = \min_{E_0}$$

где k — число нейронов сети; n — размерность признакового пространства.

Построение НС через решение задачи минимального покрытия обучающей выборки объектами-эталонами позволяет НС сохранять свойства “обобщения” и сводить к минимуму количество неверных ответов на примерах, которые она “не видела” в процессе обучения.

2. Вычислительный эксперимент

За основу модельного примера для вычислительного эксперимента взяты данные из [9, с. 307]. В состав данных входят 40 объектов (точек), описываемых двумя количественными признаками. Первые 20 точек определяют первый класс, точки с 21 по 40 — второй. В описание каждого объекта включены два номинальных признака, имеющие по две градации каждый.

Порядок предъявления объектов-кандидатов на удаление из числа эталонов процедурой “последовательное исключение” определялся исходной нумерацией (по возрастанию). Результаты вычислительного эксперимента представлены ниже.

Структура выборки	Число нейронов	Число признаков	Оценка сложности
а	9	4	36
б	10	4	40
в	11	6	66
г	5	2	10

Под литерой “г” приведены результаты, полученные при решении задачи минимального покрытия эталонами исходных данных из [9, с. 307].

Список литературы

- [1] НИКИФОРОВ А. М., ФАЗЫЛОВ Ш. Х. Методы и алгоритмы преобразования типов признаков в задачах анализа данных. Ташкент: Фан, 1988.
- [2] ГОРБАНЬ А. Н. Обучение нейронных сетей. М.: СССР – США СП “ParaGraph”, 1990.
- [3] LE CUN Y., DENKER J. S., SOLLA S. A. Optimal Brain Damage // Advances in Neural Information Proc. Systems. 1990. P. 598–605.
- [4] ДОРОГОВ А. Ю., АЛЕКСЕЕВ А. А. Пластичность многослойных нейронных сетей // Изв. вузов. Приборостроение. 1998. Т. 41, №4. С. 36–41.
- [5] ТУ ДЖ., ГОНСАЛЕС Р. Принципы распознавания образов. М.: Мир, 1978.
- [6] КОТЮКОВ В. И. Многофакторные кусочно-линейные модели. М.: Финансы и статистика, 1984.
- [7] ИГНАТЪЕВ Н. А. Распознающие системы на базе метода линейных оболочек // Автоматика и телемеханика. 2000. №3. С. 168–172.
- [8] ВАПНИК В. Н. Восстановление зависимостей по эмпирическим данным. М.: Наука. 1979.
- [9] АЙВАЗЯН С. А., БУХШТАБЕР В. М., ЕНЮКОВ И. С., МЕШАЛКИН Л. Д. Прикладная статистика: Классификация и снижение размерности. М.: Финансы и статистика. 1989.

*Поступила в редакцию 12 февраля 2001 г.,
в переработанном виде – 19 июня 2001 г.*