

ОБ ОПТИМИЗАЦИИ НЕЙРОННОЙ СЕТИ, УПРАВЛЯЕМОЙ КАТЕГОРИАЛЬНЫМИ ПАРАМЕТРАМИ¹

Ю. В. Кольцов², В. В. Гарнага³

ON OPTIMIZATION OF A NEURAL NETWORK CONTROLLED BY NOMINAL PARAMETERS

Koltsov Yu.V., Garnaga V.V.

The problem of optimization of the frame of a neural network controlled by nominal parameters is considered. The compact coding method of values of nominal variables is offered, which makes it possible to reduce the number of neurons in the input layer of a neural network. The method is illustrated on the problem of avalanches forecasting.

При решении задач, связанных с проектированием и реализацией системы прогнозирования схода снежных лавин *Avalanche* [1,2], возникла необходимость в повышении эффективности кодирования категориальных параметров. Это обусловлено тем, что некоторые параметры нейронной сети, управляющей параметрами АРПСС, являются нечисловыми. В системе *Avalanche* [1,2] к таким параметрам относятся форма верхней облачности, форма облачности среднего яруса, форма облачности нижнего яруса и вертикального развития, направление ветра. Поскольку сеть может работать только с числовыми значениями, то описанные параметры должны быть переведены в числовую форму, например, направление ветра — в угловой коэффициент. В то же время такие параметры, как форма облачности, в отличие от направления ветра нельзя задать числовым значением — они являются категориальными. Следует отметить, что количество значений категориальной переменной конечно. На практике часто применяют два метода кодирования категориальных переменных [3].

Двоичное кодирование типа « $n \rightarrow n$ ». В этом методе кодирования категориальных переменных каждому из n значений переменной ставится в соответствие набор входных значений для n бинарных нейронов. Для каждой такой переменной первое значение кодируется как $\{1, 0, 0, \dots, 0\}$, второе — как $\{0, 1, 0, \dots, 0\}$ и далее, вплоть до n -й $\{0, 0, 0, \dots, 1\}$. Недостаток этого кодирования состоит в том, что резко увеличивается количество нейронов во входном слое нейросети.

Обычно при обучении нейронной сети различные значения категориальной переменной существенно различаются числом примеров. При этом снижается информативность нейронов, отвечающих за значения категориальной переменной, которые представлены сравнительно небольшим числом примеров. То есть нейронная сеть слабо реагирует на воздействие тех значений категориальной переменной, количество которых было сравнительно мало в обучающей выборке. Это — еще один недостаток рассмотренного метода.

Двоичное кодирование типа « $n \rightarrow m$ ». В этом методе кодирования каждому из n значений категориальной переменной ставится в соответствие m входных значений бинарных нейронов, при этом число m выбирается таким образом, чтобы при обучении нейронной сети распределение активаций нейронов, кодирующих различные значения, было бы в наибольшей степени равномерно. Одним из условий достижения последнего требования является неравенство $m < n$, при этом одно категориальное значение может кодироваться входными значениями нескольких бинарных нейронов. Например, первое значение кодируется как $\{1, 1, 0, \dots, 0\}$, второе — как $\{0, 1, 1, \dots, 1\}$ и т.д. Для этого строится одномерное распределение значений категориальной переменной в обучающей выборке. С его помощью мы можем получить ответ на вопрос, как часто в выборке встречаются те или иные значения. Затем распределяем редко встречающиеся категориальные значения по нейронам, которые отвечают за более часто встречающиеся категориальные значения.

¹Работа выполнена при поддержке РФФИ Р2003ЮГ (03-01-96649).

²Кольцов Юрий Владимирович, канд. физ.-мат. наук, доцент, заведующий кафедрой информационных технологий, декан факультета прикладной математики КубГУ. E-mail: koltsov@kubsu.ru

³Гарнага Валерий Владимирович, аспирант кафедры информационных технологий факультета прикладной математики КубГУ. E-mail: valgarn@kubsu.ru

Например, рассмотрим категориальную переменную, состоящую из трех значений. Пусть в обучающей выборке первое значение встречается десять раз, второе — один раз, третье — пять раз. Тогда их можно закодировать двумя нейронами, причем первое значение кодируется как (1, 0), второе — как (1, 1), а третье — как (0, 1).

Этот метод гарантирует приблизительно одинаковую информативность всем m нейронам, кодирующим значения категориальной переменной. Недостаток этого метода заключается в добавлении нежелательной упорядоченности значениям категориальной переменной при m существенно меньше n .

В данной работе предлагается метод, который позволяет сократить число входных нейронов, отвечающих за тот или иной категориальный параметр, при этом не добавляя нежелательной упорядоченности значениям категориальной переменной.

Опишем предлагаемый метод.

Пусть существует n категориальных переменных $X_1, X_2, \dots, X_n$, область определения каждой из которых является подмножеством конечного множества $\Sigma = \{\sigma_1, \sigma_2, \dots, \sigma_k\}$. Расширим область определения каждой переменной $X_1, X_2, \dots, X_n$ до Σ .

Под соотношением значимостей на каждом категориальном значении множества Σ для переменных $X_1, X_2, \dots, X_n$ будем понимать следующую матрицу:

$$\mathbf{M} = (m_{i,j})_{i=1,n}^{j=1,k},$$

где $m_{i,j}$ выбирается исходя из практических соображений для каждого категориального значения σ_i , для категориальной переменной X_j . Формирование данной матрицы требует работы экспертов.

Тогда $X_1, X_2, \dots, X_n$ можно закодировать при помощи k нейронов следующим способом. Определим матрицу $\mathbf{A}$, размером $n \times k$:

$$\mathbf{A} = (a_{i,j})_{i=1,n}^{j=1,k},$$

Каждой j -й строке матрицы $\mathbf{A}$ ставится в соответствие значение категориальной переменной X_j , закодированное по методу двоичного кодирования типа « $n \rightarrow n$ ».

Определим вспомогательную матрицу:

$$\mathbf{B} = \mathbf{M}^T \times \mathbf{A} = (b_{i,j})_{i=1,k}^{j=1,k}.$$

Входными значениями k нейронов, которыми закодированы категориальные переменные $X_1, X_2, \dots, X_n$, являются суммы элементов соответствующей строки матрицы $\mathbf{B}$:

$$S_i = \sum_{j=1}^k b_{i,j} = \sum_{j=1}^k \sum_{l=1}^k m_{l,i} a_{l,j}, \text{ где } i = 1, 2, \dots, k.$$

Для примера рассмотрим задачу прогнозирования схода снежных лавин [1, 2]. В качестве множества Σ примем множество, состоящее из обозначений формы облачности, применяемых в специальной литературе:

$$\Sigma = \{Ac, As, Cb, Cc, Ci, Cs, Cu, Frnb, neb, Ns, Sc, Cm, St\}.$$

Заметим, что такие параметры, как форма верхней облачности, форма облачности среднего яруса, форма облачности нижнего яруса и вертикального развития, по своей сути упорядочены в отношении влияния на лавиноопасную обстановку. Обозначим это отношение как $n1:n2:n3$. На основе практических наблюдений для задачи прогнозирования схода снежных лавин выведено соотношение $n1:n2:n3 = 1:2:8$. Следуя из этого, в качестве матрицы значимостей категориальных переменных $\mathbf{M}$ выберем матрицу:

$$\mathbf{M} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ 8 & 8 & 8 & 8 & 8 & 8 & 8 & 8 & 8 & 8 & 8 & 8 & 8 \end{pmatrix}.$$

Примем во внимание тот факт, что в бинарных входных значениях нейронов 0 и 1 есть асимметрия: нулевые значения не дают никакого вклада в градиент ошибки нейронной сети, таким образом негативно влияя на процесс обучения нейронной сети [3]. В силу логической равноправности обоих значений входов установим бинарную кодировку нейронов в виде двух значений -1 и 1 , сохраняющую равноправие в процессе обучения.

Тогда на i -й нейрон входного слоя нейронной сети, который кодирует описанные категориальные переменные, должно поступать значение

$$S_i = \sum_{j=1}^{13} \sum_{l=1}^{13} m_{l,i} a_{l,j}, \quad i = 1, 2, \dots, 13.$$

Здесь $m_{p,s}$ — соответствующий элемент матрицы $\mathbf{M}$; $a_{r,t}$ — соответствующий элемент матрицы $\mathbf{A}$, которая заполняется согласно описанному методу.

Следует отметить, что матрица **A** заполняется заново на каждой обучающей выборке, поступающей на вход нейронной сети.

Поскольку каждый категориальный параметр характеризуется тринадцатью значениями, выигрыш в производительности оказывается существенным.

Предложенный метод дает возможность оптимизировать структуру нейронной сети, управляемой категориальными параметрами, так как позволяет представлять классы категориальных значений примерно одинаковым числом примеров, по сравнению с методом двоичного кодирования типа « $n \rightarrow n$ », тем не менее не прибавляя нежелательной упорядоченности категориям, как в методе двоичного кодирования типа « $n \rightarrow m$ ».

Литература

1. Кольцов Ю. В., Гарнага В. В. Распределенная информационная статистическая система Avalanche // Человек и окружающая природная среда — проблема взаимодействия: Материалы IV Междунар. науч.-практ. конф. Пенза, 2001. С. 106–110.
2. Кольцов Ю. В., Гарнага В. В. Разработка распределенной информационной статистической системы Avalanche и определение управляющих параметров // Состояние биосферы и здоровье людей: Материалы Междунар. науч.-практ. конф. Пенза, 2001. С. 105–107.
3. Ежов А., Шумский С. Нейрокомпьютинг и его применение в экономике и бизнесе. М.: МИФИ, 1998. 222 с.