

**СОВЕРШЕНСТВОВАНИЕ ПРОЦЕДУРЫ ПОДГОТОВКИ ОБУЧАЮЩИХ
МНОЖЕСТВ ДЛЯ НЕЙРОННЫХ СЕТЕЙ****Гитис В. Б., Гитис Т. П.**

Рассмотрена процедура предварительной обработки данных для обучения нейронных сетей. Проанализирована необходимость предварительной обработки данных для работы нейронных сетей. Предложено введение в нормализационную процедуру коэффициентов весовости переменных для учета особенностей решаемой задачи. Показано, что существует проблема нормализации ординальных переменных в части их внутренних уровней. Предложена к применению усовершенствованная схема нормализации, позволяющая задавать весовость как ординальной переменной в целом, так и отдельных ее уровней. Также приведены обратные нормализационным формулы для интерпретации весовых коэффициентов нейронов.

Розглянуто процедуру попередньої обробки даних для навчання нейронних мереж. Проаналізовано необхідність попередньої обробки даних для роботи нейронних мереж. Запропоновано введення в нормалізаційну процедуру коефіцієнтів ваги змінних для врахування особливостей завдання, що вирішується. Показано, що існує проблема нормалізації ординальних змінних в частині їх внутрішніх рівнів. Запропонована до застосування вдосконалена схема нормалізації, що дозволяє задавати вагу як ординальній змінній в цілому, так і окремих її рівнів. Також приведені зворотні нормалізаційним формули для інтерпретації вагових коефіцієнтів нейронів.

In the article the procedure of rough-down of information is examined for teaching of neuron networks. The necessity of rough-down of information for work of neuron networks is analyzed. Introduction to normalization procedure of coefficients of ponder ability of variable for an account features of the decided task is offered. It is shown, that the problem of normalization of ordinals of variables in part of their internal levels exists. The improved chart of normalization, allowing setting ponder ability both ordinal of variable on the whole and its separate levels, is offered to application. Reverse normalization formulas for interpretation of gravimetric coefficients of neurons are also brought.

Гитис В. Б.

канд. техн. наук, доц. кафедры ИСПР ДГМА
veniamin.gitis@dgma.donetsk.ua

Гитис Т. П.

ассистент кафедры ЭП ДГМА
tpg78@mail.ru

УДК 004.032.26

Гитис В. Б., Гитис Т. П.

СОВЕРШЕНСТВОВАНИЕ ПРОЦЕДУРЫ ПОДГОТОВКИ ОБУЧАЮЩИХ МНОЖЕСТВ ДЛЯ НЕЙРОННЫХ СЕТЕЙ

Обычно компоненты входных векторов в обучающих наборах имеют различные единицы измерения и диапазоны вариации. Часть переменных представляются непрерывными величинами, изменяющимися в ограниченном диапазоне. Другие переменные являются дискретными (например, целочисленными) и измеряются по равномерной шкале. То есть для таких переменных известно расстояние между соседними значениями, и оно постоянно (равно единице). Это обусловлено тем, что указанные переменные отражают количественные признаки, полученные путем накопления каких-либо величин.

Также существует группа переменных, полученных с помощью оцифровки качественных переменных и отражающих степень проявления некоторого качества. Этим переменным соответствует ординальная (порядковая) шкала измерения. Для такой шкалы характерно известное отношение порядка между состояниями, однако расстояние между состояниями не определено. Кодирование ординальных признаков значением одной переменной при отсутствии априорной информации о расстояниях между соседними расстояниями затруднено, поскольку такое кодирование задает эти расстояния явным образом [1].

Учет всей совокупности характеристик входного сигнала и восприятие его как единого информационного образа требует проведения определенных процедур предварительной обработки данных.

Для того чтобы представить все элементы входного сигнала числами одного типа из одного диапазона необходимо выполнить операцию масштабирования.

Подача на вход нейронной сети немасштабированных данных затрудняет обучение сети и приводит к ошибкам в ее работе, так как:

1. Работа весовых коэффициентов сети в различных масштабах усложняет начальную инициализацию весов.
2. Весовые коэффициенты нейронов в зависимости от величины дисперсии примут очень большие или очень малые значения, что увеличит время обучения и снизит точность прогнозов.
3. Нейроны входного слоя или окажутся в постоянном насыщении из-за большого усредненного значения совокупности входных данных и малой дисперсии, или будут заторможены из-за малого среднего значения выборки [2].

Также необходимо масштабирование эталонных выходных сигналов, потому что обычно диапазон выходных значений нейронов лежит в ограниченном интервале и нейронная сеть в принципе не сможет выдать ответ в реальном масштабе данных.

Поскольку нейронные сети анализируют не абсолютные значения входных сигналов, а их изменения, то для повышения различимости сигналов кроме масштабирования при предварительной обработке данных следует выполнять сдвиг входных данных. Сдвиг обеспечивается при идентификации границ диапазона изменения признака и рассмотрении их в качестве границ входного диапазона [3].

Операции сдвига и масштабирования вместе представляют собой нормализацию входных данных. Числовые сигналы необходимо масштабировать и сдвигать так, чтобы весь диапазон исходных значений попадал в нормализованный диапазон входных сигналов. В зависимости от вида активационных функций нейронов исходные сигналы нормализуются в единичный гиперкуб или в гиперкуб с координатами углов a и b по координатной оси, соответствующей масштабируемому признаку.

Целью работы является совершенствование процедуры предварительной обработки данных для обучения нейронных сетей за счет получения возможности управления весомостью переменных всех типов.

Для нормализации данных в заданный диапазон применяют минимаксное преобразование, рассчитываемое по следующей формуле [4]:

$$x_i^H = \frac{(x_i - x_{min})(b - a)}{x_{max} - x_{min}} + a, \quad (1)$$

где x_i – нормализуемая величина;

i – индекс-компонента вектора исходных данных ($i = 1 \dots n$);

n – размерность входного сигнала;

x_{min}, x_{max} – соответственно минимальное и максимальное значения нормализуемого признака;

a, b – соответственно нижняя и верхняя границы нормализованного диапазона.

При использовании симметричного относительно нуля диапазона нормализации, когда $a = -b$, и приняв $b = k$, формулу (1) можно представить в виде:

$$x_{ip}^H = \frac{2k_i(x_{ip} - x_{min\ i})}{x_{max\ i} - x_{min\ i}} - k_i, \quad (2)$$

где x_{ip}^H – нормализованное p -ое значение i -го компонента вектора исходных данных, которое будет подано на вход сети;

x_{ip} – p -ое значение i -го компонента вектора исходных данных;

p – номер примера в обучающем множестве ($p = 1 \dots P$);

P – количество примеров;

k_i – коэффициент весомости i -го фактора.

Коэффициент весомости позволяет управлять значимостью каждого входного фактора. В качестве базового значения принимается $k = 1$ для всех переменных, и тогда выходной нормализованный диапазон составит $[-1; 1]$. Увеличение этого коэффициента для некоторого фактора (или уменьшение коэффициента для остальных факторов) увеличивает расстояние между точками и позволяет подчеркнуть различие между объектами, исходя из значения рассматриваемого фактора. Управление коэффициентами весомости факторов позволяет ранжировать их роль по значимости для рассматриваемой задачи.

Применение формулы (2) решает проблему нормализации и управления весомостью непрерывных переменных. Однако ординальные переменные, должны отличаться от непрерывных по способу их кодирования и нормализации, поскольку необходимо иметь возможность управлять не только весомостью фактора в целом, но и расстояниями между уровнями ординальной переменной. Это даст возможность повышать или снижать значимость отдельных уровней критерия, исходя из особенностей моделируемого объекта.

Для получения такой возможности предлагается к применению следующая формула модификации значений уровней ординальной переменной:

$$x'_j = x_{j-1}(1 - \alpha_j) + x_{j+1}\alpha_j, \quad (3)$$

где x'_j – модифицированное значение j -го уровня ординальной переменной;

x_{j-1} и x_{j+1} – соответственно значения предыдущего и последующего уровней ординальной переменной относительно модифицируемого;

α_j – коэффициент весомости j -го уровня ординальной переменной ($\alpha \in [0; 1]$);

j – номер уровня ординальной переменной ($j = 1 \dots m$);

m – количество уровней ординальной переменной.

Равномерной шкале нормализации ординальной переменной соответствует $\alpha = 0,5$ для всех уровней. Уменьшение α_j приближает j -й уровень к $(j - 1)$ -му и, тем самым, снижает его весомость. И наоборот, увеличение α_j смещает его к $(j + 1)$ -му уровню и весомость уровня увеличивается.

Тогда, исходя из формул (2) и (3) и учитывая упорядоченность по возрастанию уровней ординальных переменных, формула для нормализации внутренних уровней примет вид:

$$x_{ip[j]}^H = \frac{2k_i \left(x_{ip[j-1]} (1 - \alpha_j) + x_{ip[j+1]} \alpha_j - x_{i[1]} \right)}{x_{i[m]} - x_{i[1]}} - k_i, \quad (4)$$

где $x_{i[1]}$ и $x_{i[m]}$ – значения, первого и последнего (m -го) уровня ординальной переменной, соответствующие минимальному и максимальному значениям этой переменной.

Нормализация первого и последнего уровней ординальной переменной осуществляется по формуле (2).

Таким образом, общая схема промежуточной минимаксной нормализации обучающего множества имеет вид:

$$\left\{ \begin{array}{l} x_{ip[j]}^H = \frac{2k_i \left(x_{ip[j]} - x_{i[1]} \right)}{x_{i[m]} - x_{i[1]}} - k_i; \forall i = \overline{1, 2}; \forall j = 1, m; \\ x_{ip[j]}^H = \frac{2k_i \left(x_{ip[j-1]} (1 - \alpha_j) + x_{ip[j+1]} \alpha_j - x_{i[1]} \right)}{x_{i[m]} - x_{i[1]}} - k_i; \forall i = \overline{1, 2}; \forall j = \overline{2, m-1}; \\ x_{ip}^H = \frac{2k_i \left(x_{ip} - x_{min i} \right)}{x_{max i} - x_{min i}} - k_i; \forall i = \overline{3, n}. \end{array} \right. \quad (5)$$

Предложенная предварительная обработка данных позволит не только корректно отображать входную информацию для нейронной сети, но и даст возможность учесть особенности моделируемого объекта.

Обратный переход от минимаксного преобразования (интерпретация) осуществляется по следующей формуле:

$$x_i = \frac{\left(x_i^H + k_i \right) \left(x_{max i} - x_{min i} \right)}{2k_i} + x_{min i}. \quad (6)$$

Формула (6) используется для непрерывных выходных сигналов, подверженных процедуре нормализации по формуле (2), а также для граничных уровней (1-го и m -го) ординальных переменных.

Для интерпретации внутренних уровней ординальных переменных необходимо вычислить текущий коэффициент ординальной переменной α^* , определяемый согласно относительной позиции полученного уровня сигнала. То есть коэффициент α^* показывает расположение рассматриваемой величины между ближайшими (меньшим и большим) нормализованными уровнями ординальной переменной. Коэффициент α^* можно рассчитать по формуле:

$$\alpha^* = \frac{x_{i[j^*]}^c - x_{i[j^*-1]}^H}{x_{i[j^*+1]}^H - x_{i[j^*-1]}^H}, \quad (7)$$

где $x_{i[j^*]}^c$ – интерпретируемое значение (выходной сигнал сети);

$x_{i[j^*-1]}^H$ и $x_{i[j^*+1]}^H$ – соответственно ближайшие к величине $x_{i[j^*]}^c$ нижнее и верхнее нормализованные значения ординальной переменной;
 j^* – условная (нецелочисленная) позиция в ординальной переменной.

Таким образом, выполняется неравенство $x_{i[j^*-1]}^H < x_{i[j^*]}^c < x_{i[j^*+1]}^H$.

Тогда интерпретация сигнала в диапазон реальных значений может быть осуществлена по формуле:

$$x_{i[j^*]} = x_{i[j^*-1]} (1 - \alpha^*) + x_{i[j^*+1]} \alpha^*, \quad (8)$$

где $x_{i[j^*-1]}$ и $x_{i[j^*+1]}$ – значения ординальной переменной в реальном масштабе данных соответствующие нормализованным позициям $x_{i[j^*-1]}^H$ и $x_{i[j^*+1]}^H$.

В результате применения формул (7) и (8) в общем случае будет получена непрерывная величина, находящаяся внутри дискретной шкалы ординальной переменной. Такое непрерывное число способно нести дополнительную информацию о полученном решении, что может быть использовано при дальнейшем анализа результата. Для практического применения необходимо округлить полученное число до ближайшей стандартизированной позиции ординальной переменной.

ВЫВОДЫ

Для обеспечения восприятия нейронными сетями разнородной информации необходимо проведение предварительной обработки данных. Введение в процедуру предварительной обработки минимаксного преобразования позволяет не только привести переменные к единому масштабу, но и дает возможность управления весомостью непрерывных переменных.

Способ нормализации ординальных переменных должен предоставлять возможность управлять не только весомостью фактора в целом, но и расстояниями между уровнями ординальной переменной. Для этого предложена формула модификации значений уровней ординальной переменной, дающая возможность повышать или снижать значимость отдельных уровней критерия, исходя из особенностей решаемой задачи.

Предложенные формулы перевода нормализованных величин в диапазон реальных данных позволяют выполнять анализ значений весовых коэффициентов нейронов, например, для определения позиций ядер в пространстве признаков.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Миркес Е. М. *Нейрокомпьютер: проект стандарта* / Е. М. Миркес; ред. В. Л. Дунин-Барковский; РАН, СО, Ин-т вычисл. моделирования. – М. : Наука: Сиб. предприятие РАН, 1999. – 190 с.
2. Заенцев И. В. *Нейронные сети: основные модели* / И. В. Заенцев. – Воронеж: ВГУ, 1999. – 157 с.
3. Горбань А. Н. *Нейроинформатика* / А. Н. Горбань, В. Л. Дунин-Барковский, А. Н. Кирдин. – Новосибирск: Наука. Сибирское предприятие РАН, 1998. – 296 с.
4. Царегородцев В. Г. *Оптимизация преобработки данных: константа Липшица обучающей выборки и свойства обученных нейронных сетей* / В. Г. Царегородцев // *Нейрокомпьютеры: разработка и применение.* – 2003. – № 7. – С. 3–8.

Статья поступила в редакцию 10.11.2011 г.