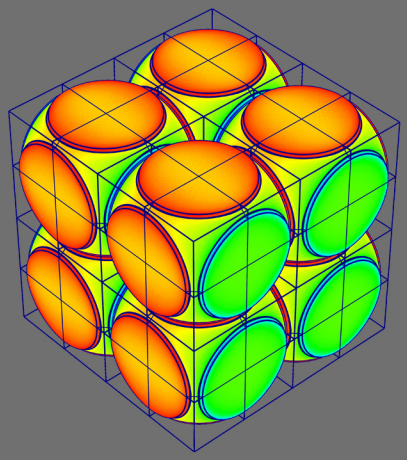
ISSN 2309-3684



Математическое моделирование и численные методы

Шовин В. А., Гольтяпин В. В. Факторное моделирование с помощью нейронной сети. Математическое моделирование и численные методы, 2016, №2 (10), с. 85-103

Источник: https://mmcm.bmstu.ru/articles/53/

Параметры загрузки:
IP: 216.73.216.140
13.11.2026 09:55:42

УДК 519.237.07

Факторное моделирование с помощью нейронной сети

© В.А. Шовин, В.В. Гольтяпин

Омский филиал ФГБУН Института математики им. С.Л. Соболева
Сибирского отделения РАН, Омск, 644099, Россия

Проведено факторное моделирование артериальной гипертензии начальной стадии с помощью метода факторизации на базе нейронной сети и алгоритма обратного распространения ошибки. Этот метод факторизации является альтернативой классическому факторному анализу. Алгоритм построения факторной структуры на базе нейронной сети был реализован программно. Данный метод был усовершенствован для проведения факторного вращения и получения интерпретируемого решения. Факторная структура артериальной гипертензии, полученная с помощью данного метода факторизации, находится в соответствии с результатами факторного моделирования посредством других методов.

Ключевые слова: *артериальная гипертензия, факторный анализ, нейронные сети, обратное распространение ошибки*

Введение. Классический факторный анализ [1] позволяет на базе выборки различных показателей сформировать факторные показатели, с необходимой точностью описывающие исходный объект и уменьшающие размерность задачи путем перехода к ним. Факторные показатели являются линейной комбинацией исходных показателей. Тем самым факторные модели носят линейный характер.

Нейронная сеть позволяет аппроксимировать отображения между исходными и целевыми показателями. При этом аппроксимируемые отображения могут иметь нелинейный характер. Двухслойный персептрон позволяет аппроксимировать любую булеву функцию булевых переменных [2]. А двухуровневая нейронная сеть способна аппроксимировать в равномерной метрике с любой заданной погрешностью $\varepsilon > 0$ любую непрерывную функцию $f(x_1, \ldots, x_n)$, а в среднеквадратической метрике — любую измеримую функцию, определенную на ограниченном множестве [3–6].

Для восстановления закономерностей между параметрами используется специальный алгоритм обучения нейронной сети — алгоритм обратного распространения ошибки [7]. Этот алгоритм с математической точки зрения представляет собой градиентный метод оптимизации.

Суть данного метода для построения факторных моделей заключается в том, что для выявления закономерностей между параметрами используется математическая модель нейронной сети с линейной передаточной функцией. Значения факторных переменных определяются равными значениями выходных сигналов нейронов скрытого

слоя нейронной сети. Тем самым нейронная сеть осуществляет классический факторный анализ, т. е. строит линейные комбинации исходных параметров [8–10].

В данной работе предлагается усовершенствованный алгоритм обратного распространения ошибки посредством введения дополнительного слагаемого в функцию ошибки для построения интерпретируемой факторной структуры и решения задачи факторного вращения на базе нейронной сети.

Математическая модель нейрона. Состояние нейрона описывается набором переменных:

- весами входных сигналов $(w_1, w_2, \ldots, w_m)$, где m — количество входных сигналов x_i;
- свободным членом w_0 в вычислении выходного сигнала. Сигнал на выходе нейрона вычисляется по формуле

$$z = \sigma(v),$$

где $v = \sum_i w_i x_i + w_0$ — взвешенная сумма сигналов на входах нейрона; σ — передаточная функция нейрона, например сигмоидальная функция $\sigma(v) = \dfrac{1}{1+\exp(-v)}$.

Нейронная сеть. Отдельные нейроны объединяются в слои. Выходные сигналы нейронов из одного слоя поступают на вход нейронам следующего слоя, модель так называемого многослойного персептрона (рис. 1). В программной реализации авторской нейронной сети вводится понятие нейронов потомков и нейронов предков. Все нейроны, имеющие входной сигнал от этого нейрона являются его потомками или пассивными нейронами или аксонами. Все нейроны, образующие входные сигналы данного нейрона, являются его предками или активными нейронами или дендритами.

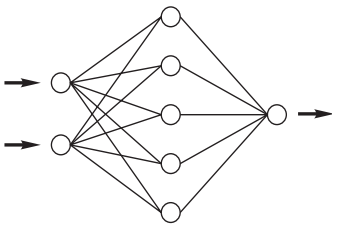


Рис. 1. Схема простой нейронной сети (входные нейроны, скрытые нейроны, выходной нейрон)

Алгоритм обратного распространения ошибки. Алгоритм обратного распространения ошибки для обучения нейронной сети соответствует минимизации функции ошибки $E(w_{ij})$. В качестве такой функции ошибки может быть использована сумма квадратов отклонений выходных сигналов сети от требуемых:

$$E = \sum_{i=1}^{m} (z_i - z_i')^2,$$

где z_i — выходное значение i-го нейрона выходного слоя; z_i' — требуемое значение i-го нейрона выходного слоя.

В данном алгоритме итерация обучения состоит из трех процедур:

1) распространение сигнала и вычисление сигналов на выходе каждого нейрона;

2) вычисление ошибки для каждого нейрона;

3) изменение весов связей.

Путем многократного цикличного подставления наборов сигналов на входе и выходе и обратного распространения ошибки производится обучение нейронной сети. Для многослойного персептрона и определенного вида передаточной функции нейрона при определенном виде функции ошибки доказана сходимость этого метода [11].

Вычисление ошибок. Если передаточная функция нейронов является сигмоидальной, то ошибки для нейронов различных слоев вычисляют по определенным формулам.

Вычисления ошибок для нейронов выходного слоя проводят по формуле

$$e_j^{(L)} = (z'_j - z_j) z_j (1 - z_j),$$

где z'_j — желаемое значение на выходе j-го нейрона выходного слоя L; z_j — сигнал на выходе j-го нейрона выходного слоя L; L — глубина нейронной сети.

Ошибки для нейронов остальных слоев рассчитывают по формуле

$$e_j^{(l)} = \left(\sum_i w_{ji}^{(l+1)} e_i^{(l+1)} \right) y_j^{(l)} (1 - y_j^{(l)}),$$

где i — индексы нейронов-потомков данного нейрона; $y_j^{(l)}$ — сигнал на выходе j-го нейрона слоя l; $w_{ji}^{(l+1)}$ — связь между j-м нейроном l-го слоя и i-м нейроном (l+1)-го слоя.

Изменение пороговых уровней нейронов и весов связей. Для изменения весов связей используют следующие формулы:

$$\Delta w_{ij}(n+1) := \eta(\alpha \Delta w_{ij}(n) + (1 - \alpha) e_j y_i),$$

$$w_{ij}(n+1) := w_{ij}(n) + \Delta w_{ij}(n+1),$$

$$\Delta w_0(n+1) := \eta(\alpha \Delta w_0(n) + (1 - \alpha) e_j),$$

$$w_0(n+1) := w_0(n) + \Delta w_0(n+1),$$

где i — индекс активного нейрона (нейрона источника входных сигналов пассивных нейронов); j — индекс пассивного нейрона; n — номер итерации обучения; α — коэффициент инерциальности для сглажива-

ния резких скачков при перемещении по поверхности целевой функции; $0 < \eta < 1$ — множитель, задающий скорость «движения».

Метод построения факторной модели. Факторный анализ основан на следующей линейной модели, связывающей исходные показатели $\vec{z_i}$ и факторы $\vec{p_i}$:

$$\vec{z_i} = a_{i1}\vec{p_1} + a_{i2}\vec{p_2} + \ldots + a_{ig}\vec{p_g} + d_i\vec{u_i},$$

где m — число переменных; g — число факторов; $\vec{z_i}$ — исходные переменные; $\vec{p_i}$ — общие факторы; $\vec{u_i}$ — специфичные факторы.

В матричном виде линейная модель факторного анализа записывается в виде

$$Z = AP + DU,$$

где $Z \underset{m\times n}{\leftrightarrow} z_{ij}$ — матрица размерности $m \times n$ значений m параметров у n объектов; $P \underset{g\times n}{\leftrightarrow} p_{ij}$ — матрица размерности $g \times n$ значений g факторов у n объектов; $U \underset{m\times n}{\leftrightarrow} u_{ij}$ — матрица размерности $m \times n$ значений m специфичных факторов у n объектов; $A \underset{m\times g}{\leftrightarrow} a_{ij}$ — матрица факторного отображения размерности $m \times g$ весовых коэффициентов; $D \underset{m\times m}{\leftrightarrow} d_{ij}$ — диагональная матрица размерности $m \times m$ весовых коэффициентов специфичных факторов.

В этом методе построения факторной модели латентные характеристики ставятся в соответствие нейронам скрытого слоя. При этом число нейронов скрытого слоя полагают меньшим числа нейронов входного слоя для осуществления факторного сжатия входной информации. Для оценки числа нейронов скрытого слоя можно применять правило классического факторного анализа Кайзера. Нейронам входного и выходного слоя ставятся в соответствие исходные характеристики объектов исследования. Когда передаточная функция нейронов линейна, такая конфигурация нейронной сети соответствует классическому факторному анализу. Тогда число нейронов входного слоя равно числу нейронов выходного слоя, число нейронов скрытого слоя меньше числа нейронов входного слоя (рис. 2).

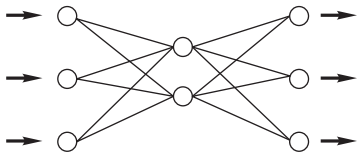


Рис. 2. Схема нейронной сети классического факторного анализа

С помощью обучения нейронной сети вычисляются веса входных связей нейронов скрытого и выходного слоя, которые соответствуют элементам обратного

и прямого факторного отображения a_{ij}. Веса нейронов ищутся в интервале [–1, 1]. Наполнение факторов исходными переменными определяется с помощью значений элементов факторного отображения и выбранного порогового уровня значимости $p \in (0,1)$. Переменная *i* входит в фактор *j*, если $|a_{ij}| \in [p,1]$.

Для раскрытия взаимосвязи факторной модели и нейронной сети применим формулы получения выходного сигнала нейронов скрытого слоя.

Обозначим выходной сигнал *j*-го нейрона скрытого слоя p_j. Выходной сигнал *i*-го нейрона входного слоя обозначим z_i. В качестве передаточной функции будем использовать линейную функцию $out = f(x) = x$.

В результате

$$p_j = f\left(\sum_{i=1}^{m} w_{ij}^{(1,2)} z_i + t_j^{(2)}\right) = \sum_{i=1}^{m} w_{ij}^{(1,2)} z_i + t_j^{(2)},$$

где m — число нейронов входного слоя; $w_{ij}^{(s,t)}$ — связь между *i*-м нейроном *s*-го слоя и *j*-м нейроном *t*-го слоя; $t_j^{(s)}$ — пороговый уровень *j*-го нейрона *s*-го слоя.

Аналогично для выходного слоя:

$$z'_j = f\left(\sum_{i=1}^{g} w_{ij}^{(2,3)} p_i + t_j^{(3)}\right) = \sum_{i=1}^{g} w_{ij}^{(2,3)} p_i + t_j^{(3)},$$

где z'_j — выходное значение *j*-го нейрона выходного слоя; g — число нейронов скрытого слоя.

Полученная линейная взаимосвязь переменных соответствует классической модели факторного анализа, в которой факторы являются линейными комбинациями исходных переменных. Задача поиска факторного отображения и значений факторов сводится к поиску весов связей и пороговых уровней нейронной сети. Поскольку факторное отображение и значения факторов являются неизвестными, необходима сеть с промежуточным слоем. Сеть в целом осуществляет тождественное преобразование, т. е. выходной сигнал на *i*-м нейроне входного слоя равен выходному сигналу *i*-го нейрона выходного слоя. Отдельные части сети (входная и выходная часть) соответствуют прямому и обратному факторному отображению.

Теорема 1. Пусть $w_{ij}^{(2,3)}$ и $w_{ij}^{(1,2)}$ — веса входных сигналов выходного и скрытого слоя нейронной сети с линейной передаточной

функцией. Число нейронов на выходном слое равно числу нейронов входного слоя. Нейронная сеть состоит из входного, скрытого и выходного слоя и осуществляет тождественное преобразование для любого входного сигнала (вектор входных сигналов сети равен вектору выходных сигналов).

Тогда выполняется следующее равенство:

$$\sum_{k=1}^{g} w_{ik}^{(1,2)} w_{ki}^{(2,3)} = 1,$$

где $w_{ij}^{(s,t)}$ — связь между i-м нейроном s-го слоя и j-м нейроном t-го слоя; g — число нейронов скрытого слоя.

Доказательство. Обозначим v_{ki} — выходной сигнал i-го нейрона k-го слоя, v_{1i} — i-й выходной сигнал нейронов первого слоя.

Для поиска весов нейронной сети необходимо выполнение условия $v_{1i} = v_{3i}$, выходной сигнал на i-м нейроне входного слоя должен быть равен выходному сигналу i-го нейрона выходного слоя. Из этого условия следует вспомогательное условие $\Delta v_{1i} = \Delta v_{3i}$, изменение i-го входного сигнала сети равно изменению i-го выходного сигнала. При этом справедливы следующие равенства:

$$v_{1i} = v_{1i}^{0} + \Delta v_{1i},$$

$$v_{3i} = v_{3i}^{0} + \Delta v_{3i},$$

где v_{1i}^{0} и v_{3i}^{0} — входной и выходной сигнал до изменения, $v_{1i}^{0} = v_{3i}^{0}$.

Допустим, что производилось изменение только i-го входного сигнала.

Из этих условий следует:

$$v_{3i} = \sum_{l=1}^{g} w_{li}^{(2,3)} v_{2l} + t_i^{(3)} = \sum_{l=1}^{g} w_{li}^{(2,3)} \left(\sum_{k=1}^{m} w_{kl}^{(1,2)} v_{1k} + t_l^{(2)} \right) + t_i^{(3)} =$$

$$= \sum_{l=1}^{g} w_{li}^{(2,3)} \left(\sum_{\substack{k=1 \\ k \neq i}}^{m} w_{kl}^{(1,2)} v_{1k} + w_{il}^{(1,2)} v_{1i} + t_i^{(2)} \right) + t_i^{(3)} =$$

$$= \sum_{l=1}^{g} w_{li}^{(2,3)} \left(\sum_{\substack{k=1 \\ k \neq i}}^{m} w_{kl}^{(1,2)} v_{1k} + w_{il}^{(1,2)} v_{1i}^{0} + w_{il}^{(1,2)} \Delta v_{1i} + t_i^{(2)} \right) + t_i^{(3)} =$$

$$= \sum_{l=1}^{g} w_{li}^{(2,3)} \left(\sum_{k=1}^{m} w_{kl}^{(1,2)} v_{1k}^{0} + w_{il}^{(1,2)} \Delta v_{1i} + t_i^{(2)} \right) + t_i^{(3)} =$$

$$= \sum_{l=1}^{g} w_{li}^{(2,3)} \left(\sum_{k=1}^{m} w_{kl}^{(1,2)} v_{1k}^{0} + t_i^{(2)} \right) + t_i^{(3)} + \sum_{l=1}^{g} w_{li}^{(2,3)} w_{il}^{(1,2)} \Delta v_{1i} =$$

$$= v_{3i}^{0} + \sum_{l=1}^{g} w_{li}^{(2,3)} w_{il}^{(1,2)} \Delta v_{1i},$$

$$\Delta v_{3i} = \sum_{l=1}^{g} w_{li}^{(2,3)} w_{il}^{(1,2)} \Delta v_{1i}.$$

Поскольку $\Delta v_{1i} = \Delta v_{3i}$, то $\sum_{l=1}^{g} w_{li}^{(2,3)} w_{il}^{(1,2)} = 1$ должно выполняться для всех i, поскольку выбор i-го входного сигнала был произволен.

Теорема доказана ■.

Веса $w_{ij}^{(2,3)}$ и $w_{ij}^{(1,2)}$ входных сигналов выходного и скрытого слоя нейронной сети с линейной передаточной функцией соответствуют коэффициентам прямого и обратного факторного отображения. Чем точнее нейронная сеть с факторным сжатием информации осуществляет тождественное преобразование, тем точнее будет выполняться равенство теоремы, соответствующее тому, что композиция прямого и обратного факторного преобразования должна давать тождественное преобразование. Докажем соответствующую теорему.

Теорема 2. Пусть $w_{ij}^{(2,3)}$ и $w_{ij}^{(1,2)}$ — веса входных сигналов выходного и скрытого слоя нейронной сети с линейной передаточной функцией. Число нейронов на выходном слое равно числу нейронов входного слоя. Нейронная сеть состоит из входного, скрытого и выходного слоя. Введем следующие величины:

Δv — средняя невязка сигнала между входом и выходом сети, приходящаяся на один входной (выходной) нейрон;

Δw — невязка равенства $\sum_{k=1}^{g} w_{ik}^{(1,2)} w_{ki}^{(2,3)} = 1$, т. е. $\Delta w = \sum_{k=1}^{g} w_{ik}^{(1,2)} w_{ki}^{(2,3)} - 1,$

где $w_{ik}^{(s,t)}$ — связь между i-м нейроном s-го слоя и k-м нейроном t-го слоя; g — число нейронов скрытого слоя.

Тогда чем меньше Δv, тем меньше Δw.

Доказательство. В предыдущей теореме доказано следующее равенство значения сигнала по одной переменной на выходе v_{3i} от приращения сигнала по той же переменной на входе Δv_{1i} :

$$v_{3i} = v_{3i}^0 + \sum_{l=1}^{g} w_{li}^{(2,3)} w_{il}^{(1,2)} \Delta v_{1i}.$$

Поскольку $\Delta v \equiv v_{3i} - v_{1i} = v_{3i} - v_{1i}^0 - \Delta v_{1i} = v_{3i} - v_{3i}^0 - \Delta v_{1i} + v_{3i}^0 - v_{1i}^0 =$ $= v_{3i} - v_{3i}^0 - \Delta v_{1i} + \Delta v^0$, где Δv^0 — начальная невязка сигналов между входом и выходом сети до изменения i-го входного сигнала, то $\Delta w = \sum_{l=1}^{g} w_{li}^{(2,3)} w_{il}^{(1,2)} - 1 = \frac{v_{3i} - v_{3i}^0}{\Delta v_{1i}} - 1 = \frac{v_{3i} - v_{3i}^0 - \Delta v_{1i}}{\Delta v_{1i}} = \frac{\Delta v - \Delta v^0}{\Delta v_{1i}}$. Это означает монотонную зависимость между Δw и Δv.

Теорема доказана ■.

Для построения нелинейных главных компонент в качестве передаточной функции может быть выбрана антисимметричная сигмоидальная функция:

$$\sigma(x) = \frac{2}{1 + \exp(-x)} - 1.$$

В любом случае независимо от вида передаточной функции для получения интерпретируемого факторного отображения вводится дополнительное слагаемое в общую целевую функцию квадратов невязок. Оно соответствует критерию «варимакс» классического факторного анализа — это максимизация дисперсии нагрузок переменной, приходящихся на все факторы:

$$\underset{i}{D} w_{ij}^{(2,3)} = \sum_{i=1}^{g} \left(w_{ij}^{(2,3)} - \underset{i}{E} w_{ij}^{(2,3)} \right)^2 \to \max,$$

$$\frac{1}{\underset{i}{D} w_{ij}^{(2,3)}} \to \min.$$

Учет «варимакс» критерия приводит к появлению дополнительных слагаемых при изменении весов нейронной сети на выходном слое:

$$\Delta w_{ij}^{(2,3)} = -\gamma \frac{\partial E}{\partial w_{ij}^{(2,3)}} = \gamma \frac{1}{\left(\underset{i}{D} w_{ij}^{(2,3)} \right)^2} \left(w_{ij}^{(2,3)} - \underset{i}{E} w_{ij}^{(2,3)} \right).$$

Другим вариантом получения интерпретируемого факторного отображения может быть использование специального критерия ин-

терпретируемости [12]. Данный критерий заключается в том, что только одна факторная нагрузка для фиксированной переменной должна быть близкой к единице, тогда как остальные должны быть близкими к нулю. Учет эмпирического критерия интерпретируемости предлагается осуществить следующим образом: среди факторных нагрузок для фиксированной переменной выбирается максимальная по модулю. Все факторные нагрузки, отличные от максимальной, уменьшаются по модулю на γ, тогда как максимальная увеличивается на γ.

Аналогичный учет главных условий факторного анализа на сумму квадратов факторных нагрузок переменной, приходящихся на все факторы, $\sum_{i=1}^{g}\left(w_{ij}^{(2,3)}\right)^2 \le 1$ и поиск факторных нагрузок в допустимом интервале $w_{ij}^{(2,3)} \in [-1,\ 1]$ приводит к поправкам на изменение весов нейронной сети для выходного слоя. В случае нарушения этих условий предлагается использовать штрафную функцию $\sum_{i=1}^{g}\left(w_{ij}^{(2,3)}\right)^2 \to \min$, соответствующую минимизации весов нейронов. Тогда $\Delta w_{ij}^{(2,3)} = -\lambda \frac{\partial E}{\partial w_{ij}^{(2,3)}} = -\lambda 2 w_{ij}^{(2,3)}$.

Для стандартизации входных значений нейронной сети используется линейное преобразование

$$y = \alpha x + \beta,$$

переводящее диапазон исходных величин x из [min, max] в [s, t].

Тогда $\alpha = \frac{t-s}{\max - \min}$, $\beta = \frac{s\max - t\min}{\max - \min}$.

Для обратного преобразования выходных значений нейронной сети из диапазона [s, t] в [min, max] используется преобразование

$$x = \frac{1}{\alpha} y - \frac{\beta}{\alpha}.$$

В качестве интервала [s, t] для антисимметричной сигмоидальной функции $\sigma(x) = \frac{2}{1+\exp(-x)} - 1$ может быть выбран интервал [–0,85, 0,85].

Тогда $\alpha = \frac{1,7}{\max - \min}$, $\beta = -0,85$.

Численный эксперимент. В качестве исходных параметров были взяты 15 биофизических показателей для 131 пациента с артериальной гипертензией начальной стадии:

1) вес;
2) индекс массы тела (ИМТ);
3) частота дыхания (ЧД);
4) сегментоядерные нейтрофилы (С);
5) лимфоциты (Л);
6) конечно-систолический размер левого желудочка (КСР);
7) конечно-систолический объем левого желудочка (КСО);
8) конечно-диастолический размер левого желудочка (КДР);
9) конечно-диастолический объем левого желудочка (КДО);
10) ударный объем (УО);
11) минутный объем сердца (МОС);
12) общее периферическое сосудистое сопротивление (ОПСС);
13) индекс Хильдебранта (ИХ);
14) фракция выброса левого желудочка (ФВ);
15) фракция укорочения левого желудочка (ФУ).

При обучении нейронной сети на данных артериальной гипертензии начальной стадии ошибка, приходящаяся на одну переменную, составляла не более 10 % от диапазона значений переменной по выборке. Эти данные содержали 131 паттерн и 15 переменных с антисимметричной сигмоидальной передаточной функцией и пятью нейронами на скрытом слое. График сходимости процесса обучения представлен на рис. 3. Под итерацией обучения понимается одна эпоха обучения, когда в сеть подставляют весь набор паттернов обучения. Под суммарной ошибкой на обучающей выборке понимается сумма ошибок для всех паттернов обучающего множества на одной итерации обучения.

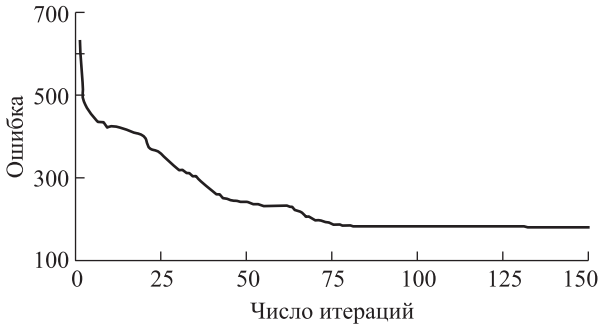


Рис. 3. График изменения суммарной ошибки
на обучающей выборке (131 паттерн, 15 переменных)

Для проверки эффективности обучения нейронной сети исходное множество входных-выходных значений было разделено на два независимых подмножества: обучающее и тестовое. Обучение проводилось на обучающем множестве, а верификация — на тестовом.

Ошибка нейронной сети на тестовом множестве является показателем того, насколько точно обучилась нейронная сеть. Относительный объем тестового множества был оценен [7] по формуле

$$r_{opt} = \frac{\sqrt{2W-1}-1}{2(W-1)},$$

где W — количество входных параметров.

При $W = 15$, $r_{opt} \approx 0{,}15$. При 131 паттерне на тестовое множество приходится 20 паттернов.

График изменения суммарной ошибки для паттернов тестового множества при верификации на каждой эпохе процесса обучения представлен на рис. 4. Под суммарной ошибкой на тестовом множестве понимается сумма ошибок для 20 паттернов тестового множества при процессе верификации на каждой эпохе обучения, т. е. когда для обучения был использован полный набор паттернов обучающего множества, но тестовое множество не участвовало в обучении. На каждой эпохе относительная ошибка для тестового множества больше относительной ошибки для обучающего множества. В пределе, когда ошибка для обучающего множества начинает сходиться, возможен эффект переобучения, т. е. значение ошибки при верификации на тестовом множестве начинает не уменьшаться, а возрастать. Это связано с тем, что промежуточные точки между точками обучающего множества в многомерном пространстве плохо аппроксимируются восстанавливаемой зависимостью нейронной сети. Данные графика изменения ошибки на тестовом множестве (рис. 4) показывают, что эффекта переобучения нет, и объем обучающего множества достаточен для количества исходных показателей, равного 15. На графике заметны лишь незначительные флуктуации ошибки на обучающем множестве в процессе сходимости ошибки для обучающего множества.

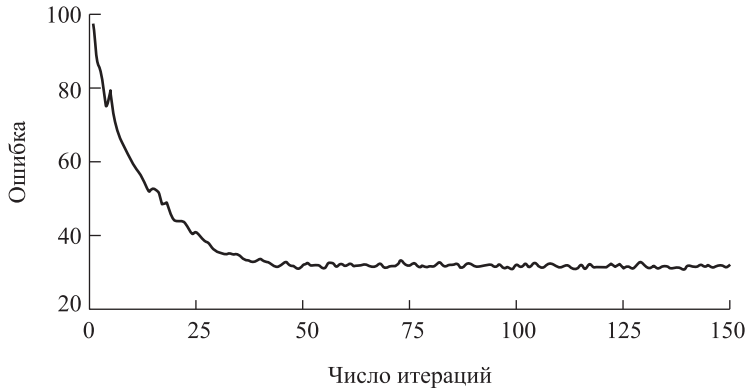


Рис. 4. График изменения суммарной ошибки
на тестовом множестве (20 паттернов, 15 переменных)

Из данных графика видно, что эффекта переобучения не наблюдается, дальнейшее обучение приводит лишь к небольшой флуктуации суммарной ошибки на тестовом множестве.

Средние ошибки, приходящиеся на 15 переменных для одного паттерна на обучающем и тестовом множестве, равны 1,28 и 1,54. При исходном диапазоне изменения параметров [–0,85, 0,85] ошибки, приходящиеся на одну переменную, для обучающего и тестового множества равны 5 и 6 %. Например, для параметра «вес» наибольшее значение было равно 116 кг, наименьшее — 45 кг, при диапазоне в 71 кг, ошибка в 6% соответствует 4,26 кг. Это свидетельствует о хорошей способности нейронной сети к обобщению. Ошибка в 6 % для 15 входных параметров и 131 примера для обучения меньше теоретической оценки ошибки в 10 %. Когда требуется 15–10 примеров для обучения, можно говорить о достаточности обучающего множества. Известно, что между действительно достаточным размером множества обучения и теоретическими оценками может существовать большой разрыв [7].

Число нейронов на скрытом слое было выбрано согласно правилу факторного анализа Кайзера, когда число факторов должно быть не больше количества собственных значений корреляционной матрицы переменных больше единицы. График собственных значений исходных переменных приведен на рис. 5.

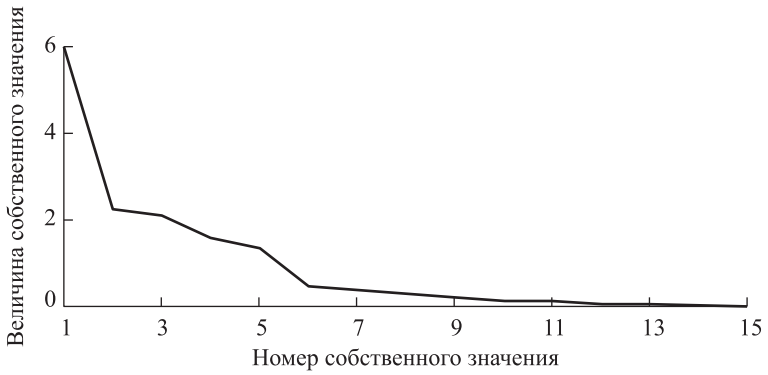


Рис. 5. Собственные значения исходных переменных

В ходе численного эксперимента с данными артериальной гипертензии было установлено, что хорошая обучаемость нейронной сети достигается при числе нейронов в скрытом слое, который не меньше верхней границы числа собственных значений исходных переменных по правилу Кайзера. При меньшем числе нейронов в скрытом слое наблюдались значительные ошибки на обучающей выборке, и наоборот: чем больше нейронов выбиралось на скрытом слое, тем меньше была ошибка на тестовом и обучающем множестве. Это связано с факторным сжатием и потерей информации на скрытом слое.

Факторное отображение, полученное с помощью нейронной сети, незначительно отличается от факторного отображения, получаемого методом главных компонент с последующим «варимакс» вращением, и имеет тот же интерпретационный характер. Все значимые факторные нагрузки двух матриц факторных отображений близки друг к другу. Соответствующие факторные структуры приведены в табл. 1–3.

С точки зрения совпадения факторных структур ошибка прогноза входного паттерна менее 10 % оказалась некритичной, и нейронная сеть является достаточно обученной для проведения факторного анализа. Данные факторные структуры подтверждаются работой [11].

Таблица 1

Факторная структура «варимакс»
(метод главных компонент + «варимакс» вращение)

Показатели	F_1	F_2	F_3	F_4	F_5
Вес	0,135	0,078	–0,073	–0,07	**0,861**
ИМТ	0,085	0,095	–0,147	–0,07	**0,858**
ЧД	0,087	0,043	–0,078	**–0,849**	0,087
С	–0,07	**–0,913**	–0,051	0,06	–0,108
Л	0,012	**0,923**	0,029	–0,017	0,065
КСР	**0,739**	0,015	–0,637	–0,093	0,151
КСО	**0,73**	–0,005	–0,628	–0,083	0,137
КДР	**0,943**	0,048	–0,256	–0,011	0,119
КДО	**0,932**	0,016	–0,299	–0,034	0,132
УО	**0,947**	0,053	0,07	–0,001	0,107
МОС	**0,93**	0,017	0,046	0,011	0,039
ОПСС	**–0,776**	–0,022	–0,117	0,072	0,027
ИХ	0,005	–0,028	0,039	**0,858**	–0,044
ФВ	–0,025	0,005	**0,845**	0,024	–0,113
ФУ	–0,028	0,072	**0,743**	0,062	–0,052

Таблица 2

Факторная структура на базе нейронной сети («варимакс» критерий)

Показатели	F_1	F_2	F_3	F_4	F_5
Вес	0,483	–0,116	–0,024	0,127	**0,846**
ИМТ	0,409	–0,062	–0,165	0,132	**0,888**
ЧД	–0,27	–0,442	–0,071	**0,852**	–0,019
С	0,02	**0,876**	0,18	0,393	–0,269
Л	–0,08	**–0,786**	–0,276	–0,503	0,214
КСР	**0,759**	–0,155	–0,575	0,217	–0,169

Окончание табл. 2

Показатели	F_1	F_2	F_3	F_4	F_5
КСО	**0,805**	–0,137	–0,525	0,187	–0,193
КДР	**0,864**	–0,296	–0,145	0,098	–0,237
КДО	**0,899**	–0,288	–0,176	0,157	–0,234
УО	**0,838**	–0,402	0,251	0,077	–0,261
МОС	**0,835**	–0,331	0,242	0,061	–0,352
ОПСС	**–0,697**	0,441	–0,401	–0,197	0,44
ИХ	0,287	0,35	–0,015	**–0,596**	–0,061
ФВ	–0,192	–0,22	**0,955**	–0,08	–0,056
ФУ	–0,152	–0,175	**0,973**	–0,212	–0,076

Таблица 3

Факторная структура на базе нейронной сети (критерий интерпретируемости)

Показатели	F_1	F_2	F_3	F_4	F_5
Вес	–0,01	0	–0,017	0,006	**0,996**
ИМТ	0,002	–0,011	0,015	–0,002	**0,989**
ЧД	–0,001	0,035	0,012	**–0,996**	0,016
С	–0,013	**–0,997**	0,016	0,006	0,004
Л	0,001	**1,000**	–0,006	–0,005	0,001
КСР	**0,869**	0,002	0,49	–0,009	0,015
КСО	**0,864**	–0,02	0,502	–0,037	0,009
КДР	**0,99**	0,005	0,004	0,009	0,004
КДО	**0,994**	0,002	0,081	0,005	–0,002
УО	**0,964**	0,004	–0,253	0,024	–0,02
МОС	**0,974**	0,003	–0,211	–0,016	–0,01
ОПСС	**–0,985**	0,008	0,267	0,035	0,113
ИХ	0,036	0,024	–0,001	**0,817**	–0,023
ФВ	0,001	–0,018	**–0,994**	–0,026	0,002
ФУ	–0,01	0,009	**–0,99**	0,034	0,02

Специальный критерий интерпретируемости оказался более эффективным, чем «варимакс» критерий. Разделение исходных переменных на факторы более мощное: незначимые факторные нагрузки близки к нулю, тогда как значимые близки к единице. Незначимые факторные нагрузки отдают свою долю дисперсии переменных (приходятся на факторы) значимым факторным нагрузкам, тем самым повышая эффективность разбиения исходных параметров на факторы. На конечном этапе факторного моделирования осуществляется интерпретация наполнения факторов переменными и более эффектив-

ное разбиение исходных переменных на факторы, соответствующее целям классического факторного вращения. Это является плюсом факторного моделирования на базе нейронной сети со специальным критерием интерпретируемости. Расхождение величин незначимых факторных нагрузок матриц факторного отображения до и после факторного вращения является обычным эффектом. Его целью является повышение эффективности интерпретационного разделения исходных переменных на факторы.

Матрица корреляций факторов немного отличается от ортогональной, что соответствует общему косоугольному факторному анализу (табл. 4).

Таблица 4

Корреляции факторов, полученных на базе нейронной сети с использованием критерия интерпретируемости

Факторы	F_1	F_2	F_3	F_4	F_5
F_1	1,000	0,089	0,146	–0,108	0,217
F_2	0,089	1,000	–0,057	–0,065	0,157
F_3	0,146	–0,057	1,000	–0,149	0,217
F_4	–0,108	–0,065	–0,149	1,000	–0,143
F_5	0,217	0,157	0,217	–0,143	1,000

Факторное моделирование позволяет выделить группы взаимосвязанных параметров, формирующих факторы заболевания артериальной гипертензии. Факторы проинтерпретированы в ранних работах [13, 14]. Рассмотрим некоторые из них.

Гемодинамический фактор включает параметры, описывающие центральную и периферическую гемодинамику, и его можно назвать главным. Переменные УО, МОС, ОПСС определяют уровень артериального давления. В норме изменениям минутного объема циркуляции должна соответствовать адекватная по величине и направлению реакция прекапиллярного русла, которая бы нивелировала эти изменения и сохраняла среднее давление на нормальном уровне. Например, если МОС снижен, то артериолы должны сузиться. Если МОС увеличен, то артериолы должны расшириться. Нарушения взаимосвязи этих показателей лежат в основе изменений уровня артериального давления. Вместе с тем изменение уровня артериального давления взаимосвязано с модуляцией сердца, за которую отвечают параметры КСР, КСО, КДР, КДО.

Фактор, составленный из параметров «фракция выброса левого желудочка» и «фракция укорочения левого желудочка». Его можно считать важным для непосредственной оценки контрактильной (сократительной, нагнетательной) функции левого желудочка. Этот

фактор определяет объемную ресурсоемкость левого желудочка. Он показывает, насколько использованы объемные резервы самого сердца для поддержания уровня артериального давления.

Кроме того, были выделены *фактор, отвечающий за соответствие массы и роста*, и *фактор, характеризующий уровень слаженности работы сердца и легких* (он определяется через частоту дыхания и индекс Хильдебранта).

Также был рассмотрен *иммунологический фактор*, который может отражать психосоматическое состояние индивида, поскольку этот фактор активируется в стрессовых состояниях. Основной вклад в формирование этого фактора вносят сегментоядерные нейтрофилы и лимфоциты.

Выделенные факторы являются различными аспектами заболевания. Например, такой фактор риска, как ожирение, соответствует нарушению фактора, отвечающего за соответствие массы и роста. Иммунологический и фактор, характеризующий уровень слаженности работы сердца и легких, могут быть объединены в один общий фактор, соответствующий стрессовой восприимчивости. Все выделенные факторы подтверждены независимыми медицинскими исследованиями.

В ранней работе [13] даны рекомендации по нормализации выделенных факторов. Эти пять факторов позволяют указать группу параметров, на которую нужно воздействовать, чтобы получить максимальный эффект от лечения. Например, для стабилизации уровня артериального давления следует воздействовать на всю группу признаков, описывающих гемодинамический фактор. При этом следует учитывать ремоделирование сердца (структурно-геометрическое состояние) при формировании патофизиологических взаимоотношений в системе кровообращения у пациентов с гипертонической болезнью. Поскольку ожирение является одним из факторов риска, то снижение веса позволит нормализовать фактор, отвечающий за соответствие массы и роста. Исключение стрессовых ситуаций пациентом позволит улучшить показатели, формирующие иммунологический фактор, а также нормализовать фактор, характеризующий уровень слаженности работы сердца и легких.

Заключение. Рассмотрен известный альтернативный метод построения факторной модели на основе нейронной сети и алгоритма обратного распространения ошибки. Данный метод был усовершенствован для проведения факторного вращения и получения интерпретируемого решения. Преимущество данного метода заключается в том, что он объединяет в себе все этапы классического факторного анализа: поиск факторного отображения, факторное вращение и вычисление значений факторов. Этот метод осуществляет косоугольный факторный анализ, тем самым имеет максимальную степень общности для линейной модели.

На базе нейронной сети с нелинейной передаточной функцией получен вариант нелинейного факторного анализа. Факторная структура артериальной гипертензии, полученная с помощью нейронной сети и критериев интерпретируемости, незначительно отличается от факторной структуры, полученной методом главных компонент с последующим «варимакс» вращением, и имеет тот же интерпретационный характер.

Представляется актуальным улучшить алгоритм факторного вращения на базе нейронной сети с помощью методов распараллеливания [15].

ЛИТЕРАТУРА

[1] Иберла К. *Факторный анализ.* Пер. с нем. В.М. Ивановой; предисл. А.М. Дуброва. Москва, Статистика, 1980.

[2] Гаврилкевич М. Введение в нейроматематику. *Обозрение прикладной и промышленной математики*, 1994, т. 1, вып. 3.

[3] Hornik K., Stinchcombe M., White H. Multilayer Feedforward Networks are Universal Approximators. *Neural Networks*, 1989, vol. 2, no. 5, pp. 359–366.

[4] Cybenko G. Approximation by Superpositions of a Sigmoidal Function. *Mathematics of Control*, *Signals and Systems*, 1989, vol. 2, pp. 303–314.

[5] Funahashi K. On the Approximate Realization of Continuous Mappings by Neural Networks. *Neural Networks*, 1989, vol. 2, no. 3, issue 4.

[6] Горбань А.Н. Обобщенная аппроксимационная теорема и вычислительные возможности нейронных сетей. *Сибирский журнал вычислительной математики*, 1998, т. 1, № 1, с. 11–24.

[7] Хайкин С. *Нейронные сети: Полный курс.* 2-е изд., испр. Москва, Вильямс, 2008, 1103 с.

[8] Осовский С. *Нейронные сети для обработки информации*. Москва, Финансы и статистика, 2002, 344 с.

[9] Gorban A., Kegl B., Wunsch D., Zinovyev A. *Principal Manifolds for Data Visualisation and Dimension Reduction*. Berlin, Heidelberg, New York, Springer, 2007.

[10] Kruger U., Antory D., Hahn J., Irwin G.W., McCullough G. Introduction of a nonlinearity measure for principal component models. *Computers&Chemical Engineering*, 2005, no. 29 (11), pp. 2355–2362.

[11] Jain A.K., Mao J., Mohiuddin K.M. Artificial Neural Networks: A Tutorial. *Computer*, 1996, vol. 29, no. 3, pp. 31–44.

[12] Шовин В.А., Гольтяпин В.В. Методы вращения факторных структур. *Математические структуры и моделирование*, 2015, № 2, с. 75–84.

[13] Гольтяпин В.В., Шовин В.А. Косоугольная факторная модель артериальной гипертензии первой стадии. *Вестник Омского университета*, 2010, № 4, с. 120–128.

[14] Шовин В.А. Конфирматорная факторная модель артериальной гипертензии. *Компьютерные исследования и моделирование*, 2012, т. 4, № 4, с. 885–894.

[15] Мартыненко С.И. О построении параллельных многосеточных алгоритмов. *Математическое моделирование и численные методы*, 2015, № 2 (6), с. 105–120.

Статья поступила в редакцию 18.05.2016

Ссылку на эту статью просим оформлять следующим образом:

Шовин В.А., Гольтяпин В.В. Факторное моделирование с помощью нейронной сети. *Математическое моделирование и численные методы*, 2016, № 2(10), с. 85–103.

Шовин Владимир Алексеевич — научный сотрудник Омского филиала Федерального государственного бюджетного учреждения науки Института математики им. С.Л. Соболева Сибирского отделения РАН. e-mail: v.shovin@mail.ru

Гольтяпин Виктор Викторович — канд. физ.-мат. наук, доцент, старший научный сотрудник Омского филиала Федерального государственного бюджетного учреждения науки Института математики им. С.Л. Соболева Сибирского отделения РАН. e-mail: goltyapin@mail.ru

Factorial modeling using neural network

© V.A. Shovin, V.V. Goltyapin

Sobolev Institute of Mathematics, Omsk branch,
Siberian Branch of the Russian Academy of Sciences, Omsk, 644099, Russia

The paper deals with the factorial modeling of the initial stage arterial hypertension. The modeling was carried out by the factorization method based on the neural network and the back propagation of error algorithm. This factorization method is an alternative to the classical factor analysis. We implemented an algorithm for constructing the factorial structure based on the neural network in software. This method has been improved for the factor rotation and obtaining an interpretable solution. The hypertension factorial structure obtained by this factorization method is in accordance with the results of the factorial modeling by other methods.

Keywords: *arterial hypertension, factor analysis, neural networks, back propagation of error.*

REFERENCES

[1] Iberla K. *Faktornyy analiz* [Factor analysis]. Transl. from German by V.M. Ivanova; Preface A.M. Dubrov. Moscow, Statistika Publ., 1980.

[2] Gavrilkevich M. Vvedenie v neiromatematiku [Introduction to neuro mathematics]. *Obozrenie prikladnoy i promyshlennoy matematiki*. *Review of applied and industrial mathematics*, vol. 1, iss. 3.

[3] Hornik K., Stinchcombe M., White H. *Neural Networks*, 1989, vol. 2, no. 5, pp. 359–366.

[4] Cybenko G. *Mathematics of Control*, *Signals and Systems,* 1989, vol. 2, pp. 303–314.

[5] Funahashi K. *Neural Networks*, 1989, vol. 2, no. 3, iss. 4.

[6] Gorban A.N. *Sibirskiy zhurnal vychislitelnoy matematiki — Siberian Journal of Numerical Mathematics*, 1998, vol. 1, no. 1, pp. 11–24.

[7] Haykin S. *Neironnye seti: Polnyy kurs* [Neural networks: A Comprehensive Foundation]. 2nd ed., corr. Moscow, Vilyams Publ., 2008, 1103 p.

[8] Osovskiy S. *Neironnye seti dlya obrabotki informatsii* [Neural networks for information processing]. Moscow, Finansy i statistika Publ., 2002, 344 p.

[9] Gorban A., Kegl B., Wunsch D., Zinovyev A., *Principal Manifolds for Data Visualisation and Dimension Reduction*. Berlin, Heidelberg, New York, Springer, 2007.
[10] Kruger U., Antory D., Hahn J., Irwin G.W., McCullough G. *Computers & Chemical Engineering*, 2005, no. 29 (11), pp. 2355–2362.
[11] Jain A.K., Mao J., Mohiuddin K.M. *Computer*, 1996, vol. 29, no. 3, pp. 31–44.
[12] Shovin V.A., Goltyapin V.V. *Matematicheskie struktury i modelirovanie — Mathematical Structures and Modeling*, 2015, no. 2, pp. 75–84.
[13] Goltyapin V.V., Shovin V.A. *Vestnik Omskogo universiteta — Herald of Omsk University*, 2010, no. 4, pp. 120–128.
[14] Shovin V.A. *Kompyuternye issledovaniya i modelirovanie — Computer Research and Modeling*, 2012, vol. 4, no. 4, pp. 885–894.
[15] Martynenko S.I. *Matematicheskoe modelirovanie i chislennye metody — Mathematical Modeling and Computational Methods*, 2015, no. 2 (6), pp. 105–120.

Shovin V.A., Research Scientist of Sobolev Institute of Mathematics, Omsk branch, Siberian Branch of the Russian Academy of Sciences. e-mail: v.shovin@mail.ru

Goltyapin V.V., Cand. Sci. (Phys. & Math.), Assoc. Professor, Senior Research Scientist of Sobolev Institute of Mathematics, Omsk branch, Siberian Branch of the Russian Academy of Sciences. e-mail: goltyapin@mail.ru