

ПРОГНОЗ МЕТОДОМ ЭКСТРАПОЛИРОВАНИЯ ОШИБКИ НЕЙРОСЕТИ

FORECASTING BY EXTRAPOLATION
OF A NEURAL NETWORK ERRORA. Gusev
A. Okunev

Summary. In this paper, we describe a method for extrapolating a neural network error developed by the authors. The method is effective when observations are made for several time periods and for several territories (objects). The method of extrapolating the error of a neural network is based on the procedure for compressing the information space and the procedure for expanding the information space. The authors considered three methods of compressing the information space: compression by all indicators, by defining indicators and by the predicted indicator. The article considers four ways of extrapolating the error of a neural network. On a concrete example in work it is shown, that the method of extrapolation of a neural network error allows to reduce essentially the general average error of the forecast.

Keywords: indicator forecast, information space, neural network, compression procedure, extension procedure.

Гусев Андрей Леонидович

Д.т.н., к.ф.-м.н., профессор, Пермский
государственный национальный исследовательский
университет
alguseval@mail.ru

Окунев Александр Анатольевич

Аспирант, Пермский государственный национальный
исследовательский университет
alexander2510@mail.ru

Аннотация. В настоящей статье описан метод экстраполирования ошибки нейронной сети, разработанный авторами. Метод представляется действенным, когда наблюдения ведутся в течение нескольких временных периодов и для нескольких территорий (объектах). Метод экстраполирования ошибки нейронной сети основан на процедуре сжатия информационного пространства и процедуре расширения информационного пространства. Авторами рассмотрены три метода сжатия информационного пространства: сжатие по всем показателям, по определяющим показателям и по прогнозируемому показателю. В статье рассмотрены четыре способа экстраполяции ошибки нейронной сети. На конкретном примере в работе показано, что метод экстраполирования ошибки нейронной сети позволяет существенно уменьшить общую среднюю ошибку прогноза.

Ключевые слова: прогноз показателя, информационное пространство, нейронная сеть, процедура сжатия, процедура расширения.

В последнее время решение задач прогнозирования тех или иных показателей в различных предметных областях становится актуальным, и поэтому появляются всё новые и новые методы и алгоритмы прогнозирования. Достаточно привести лишь несколько научных работ [1–8], чтобы убедиться в этом.

Авторы в своей статье [9] предлагали для решения задачи прогнозирования метод экстраполирования ошибки нейронной сети на основе процедур сжатия-расширения информационного пространства. Общую задачу прогнозирования авторы интерпретировали следующим образом. Прогнозируемый показатель y может быть спрогнозирован по известному определяющему вектору $X = (x_1, x_2, \dots, x_m)$ и определяющему вектору $Z = (z_1, z_2, \dots, z_n)$, который является неизвестным. То есть вектор X и вектор Z несут информационную суть прогноза показателя y . Следуя методу экстраполирования ошибки нейросети, можно оценить воздействие определяющего вектора Z на показатель y , как скалярного показателя z . Со времени опубликования статьи [9], авторы развили этот метод и в настоящей работе приводят новые результаты.

Метод экстраполирования
ошибки нейронной сети

Пусть решается задача прогноза какого-либо показателя в зависимости от определяющих показателей. Для этого фиксируются все перечисленные показатели на протяжении T периодов по условным L территориям (объектам) с необходимым лагом. Таким образом, получается $M = T * L$ наблюдений. Из этого множества наблюдений, по определенным правилам, отбираются наблюдения, для которых можно спроектировать и построить нейронную сеть с достаточно простой структурой, которая будет иметь ошибку близкую к нулю.

Правила отбора наблюдений могут быть реализованы пошаговой итерационной процедурой и состоять в следующем.

Шаг 1. Для каждого показателя найдем среднее значение и стандартное отклонение по M наблюдениям. Далее помечаем те наблюдения, для которых не выполняется хотя бы один раз следующие условия:

$$m_{x_i} - k\sigma_{x_i} \leq x_{i,j} \leq m_{x_i} + k\sigma_{x_i},$$

$$m_y - k\sigma_y \leq y_j \leq m_y + k\sigma_y,$$

где $i = (\overline{1, m})$, $j = (\overline{1, M})$, m_∞ — среднее значение, σ_\bullet — стандартное отклонение, k — некоторый наперед заданный коэффициент (например, $k = 5$ или $k = 10$), $x_{i,j}$ — j -е наблюдение i -ого определяющего показателя, y_j — j -е наблюдение прогнозируемого показателя.

Шаг 2. На следующем шаге исключим из дальнейшего рассмотрения помеченные наблюдения. После этого останется M_I наблюдений ($M_I \leq M$).

Шаг 3. На сжатом информационном пространстве пытаемся построить нейросетевую модель прогноза. В случае если результаты не удовлетворяют наперед заданным условиям (например, среднеквадратичная ошибка прогноза нейросетевого моделирования должна быть меньше наперед заданного числа ε_0 близкого к нулю), то возвращаемся к первому шагу процедуры. При этом выбираем новое значение коэффициента k , обозначим его k_I ($k_I \leq k$). Желательно, чтобы коэффициент k_I не сильно отличался от k (например, $k_I = k$ или $k_I = k - 0,01$). Далее переходим к первому шагу итерационной процедуры со значением $M = M_I$.

Следуя изложенной пошаговой итерационной процедуре сжатия информационного пространства, часто можно добиться вполне удовлетворительных результатов (например, среднеквадратичная ошибка нейросетевого моделирования меньше наперед заданного числа ε_0 близкого к нулю), тогда как без ее применения построить сколько-нибудь приемлемую нейросетевую модель невозможно.

Если удаётся построить на сжатом информационном пространстве удовлетворительную нейросетевую модель, то далее в неё загружаются все M наблюдений, и вычисляется ошибка (погрешность) для каждого из M наблюдений. Ошибка для каждого наблюдения является дополнительным скалярным показателем z (условным показателем, который отображает общее воздействие неучтенных в нейросети показателей, влияющих на прогнозируемый показатель — это расширение информационного пространства).

Здесь необходимо заметить и подчеркнуть, что на реальных примерах авторы убедились в том, что для различных практических случаев сжатие целесообразно проводить по разным сценариям: по всем показателям; только по определяющим показателям; только по прогнозируемому показателю.

Далее необходимо рассчитать показатель z по каждой территории для прогнозируемого временного пе-

риода. По каждой территории (объекту) имеем T значений показателя z по временным периодам: z_1, z_2, \dots, z_T . Необходимо определить z_{T+1} .

Способы экстраполяции ошибки нейросети

1. Самый простой способ — использовать в качестве z_{T+1} на прогнозируемый период значение z_T за последний исследованный временной период по каждой территории (объекту), то есть $z_{T+1} = z_T$.

2. Для прогноза следует использовать среднее значение за все временные периоды

$$z_{T+1} = \sum_{i=1}^T z_i / T$$

по каждой территории (объекту).

3. Для прогноза необходимо проанализировать значения z_1, z_2, \dots, z_T и вычислить z_{T+1} , то есть экстраполировать z на прогнозируемый временной период по каждой территории (объекту). Может сложиться три ситуации.

Первая ситуация. Значения z постоянно возрастают (не убывают) от временного периода к временному периоду. Тогда нужно рассчитать среднее значение возрастания z за один период

$$(\Delta z = \sum_{i=1}^{T-1} (z_{i+1} - z_i) / (T - 1)).$$

Для прогноза следует использовать значение z за последний исследованный временной период плюс среднее значение возрастания z за один временной период ($z_{T+1} = z_T + \Delta z$). Дополнительно, можно исследовать возрастание значения z в течение времени ($\Delta z_i = z_{i+1} - z_i$, $i = \overline{1, T-1}$). Например, если с течением времени Δz_i увеличивается в среднем на величину $\Delta \Delta z_i$, то необходимо вычислить среднее увеличение за все исследуемые временные периоды и в качестве z_{T+1} для прогноза брать сумму: $z_{T+1} = z_T + \Delta z + \Delta \Delta z$.

Вторая ситуация. Значение z постоянно убывает (не возрастает) от временного периода к временному периоду. Тогда нужно рассчитать среднее значение убывания за один период

$$(\Delta z = \sum_{i=1}^{T-1} (z_i - z_{i+1}) / (T - 1)).$$

Для прогноза следует использовать значение z за последний исследованный временной период минус среднее значение убывания z за один временной период ($z_{T+1} = z_T - \Delta z$). Дополнительно, можно исследовать

убывание значения z в течение времени ($\Delta z_i = z_i - z_{i+1}$, $i = \overline{1, T-1}$). Если с течением времени Δz_i уменьшается в среднем на величину $\Delta \Delta z$, то необходимо вычислить среднее уменьшение z за все исследуемые временные периоды и в качестве z_{T+1} для прогноза брать разность: $z_{T+1} = z_T - \Delta z - \Delta \Delta z$.

Третья ситуация. Значение z то возрастает, то убывает от временного периода к временному периоду. Тогда для прогноза следует использовать значение z за последний исследованный временной период ($z_{T+1} = z_T$) или среднее значение за все периоды

$$(z_{T+1} = \sum_{i=1}^T z_i / T).$$

Дополнительно, можно исследовать возрастание и убывание значения z в течение времени. В зависимости от большего числа возражений или убываний z за все временные периоды следует вычислять z в соответствии с первой или второй ситуацией.

Применить процедуру сглаживания для значений z_1, z_2, \dots, z_T . Это означает, что над последовательностью, которая, то возрастает, то убывает, выполняется преобразование, в результате которого она становится монотонной.

В статье [1] сглаживание с помощью двустороннего скользящего среднего применяется для удаления шумовой компоненты и выделения тренда временного ряда. Вычисление скользящего среднего выполняется следующим образом:

$$\begin{cases} z'_t = \frac{1}{2k-1} \sum_{i=t-k+1}^{t+k-1} z_i \text{ npu } t = k, \dots, T-k+1 \\ z'_t = \frac{1}{t+k-1} \sum_{i=1}^{t+k-1} z_i \text{ npu } t = 1, \dots, k-1 \\ z'_t = \frac{1}{T+k-t} \sum_{i=t-k+1}^T z_i \text{ npu } t = T-k+2, \dots, T \end{cases}$$

где k — параметр сглаживания, z — последовательность, в которой необходимо выделить тренд, z' — монотонная последовательность.

Множественное применение вышеописанного преобразования к последовательности z приводит к устранению всех отклонений от возрастающего или убывающего тренда. Как только получена монотонная последовательность, можно вычислить z способом, описанным в предыдущем пункте.

Данный способ имеет смысл, так как значения z в последовательности часто отклоняются от тренда и без применения сглаживания пункт 3 по сути сводится к пункту 2.

Применив один из способов экстраполяции ошибки нейросети, теперь достаточно подставить значения вектора X для прогнозируемого временного периода по территориям (объектам) в построенную нейронную сеть. Таким образом, получим псевдо прогноз по территориям (объектам). Для получения истинного прогноза по территориям (объектам) к псевдо прогнозу необходимо прибавить рассчитанное значение Z_{T+1} для прогнозируемого временного периода по территориям (объектам). Эта сумма и будет значением прогнозируемого показателя по территориям (объектам).

Замечание

В начале статьи авторы предполагают, что прогнозируется один показатель. Как известно, нейросетевые технологии позволяют прогнозировать одновременно несколько показателей. Встает вопрос. Следует ли применять метод сжатия-расширения информационного пространства при прогнозе нескольких показателей.

Экспериментируя с методом сжатия-расширения информационного пространства при прогнозе, авторы пришли к следующему выводу. Если прогнозируемые показатели имеют между собой естественную связь, то в результате применения метода сжатия-расширения информационного пространства и построения нейросетевой модели прогноза для всех показателей одновременно, общая средняя ошибка прогноза всех показателей будет меньше, чем средняя ошибка моделей для каждого отдельного показателя. Если же прогнозируемые показатели не имеют между собой естественной связи, то в результате применения метода сжатия-расширения информационного пространства и построения нейросетевой модели прогноза для всех показателей одновременно, общая средняя ошибка прогноза всех показателей будет больше, чем средняя ошибка моделей для каждого отдельного показателя. В качестве показателей, имеющих естественную связь, можно рассматривать, например, показатели, сумма которых равна 100%. Например, при прогнозе процентной возрастной смертности по 46 территориям, авторам удалось спрогнозировать смертность по 17 возрастным группам (то есть одновременно прогнозировались 782 показателя) с общей средней ошибкой 1,28%.

Пример

Для проверки метода прогнозирования из статистического справочника [10] был случайным образом выбран показатель «Болезни крови, кроветворных органов и отдельные нарушения, вовлекающие иммунный механизм». В справочнике содержатся 5 показателей качества среды обитания и 12 (выбрано 4) показателей ка-

чества медицинского обслуживания. Все показатели анализировались за 10 лет (2005–2014 годы). В справочнике оказалось 77 территорий пригодных для статистической обработки данных. Задача заключалась в том, чтобы по показателям качества среды обитания и качества медицинского обслуживания предыдущего года спрогнозировать показатель заболеваемости следующего года. Следовательно, после сдвига статистических данных с лагом в один год оказалось, что нейросетевую модель прогноза можно построить по 8 годам (временным периодам) и проверить её качество (ошибка прогноза должна быть не более 3% от размаха распределения показателя заболеваемости) по 2014 году. Таким образом, для обучения нейронной сети были в распоряжении $M = T \cdot L = 8 \cdot 77 = 616$ наблюдений.

Оговоримся сразу, что под ошибкой прогноза будем подразумевать среднюю ошибку по территориям. А под ошибкой прогноза по территории будем подразумевать модуль разности между прогнозным и фактическим показателями заболеваемости по территории, разделенный на размах показателя заболеваемости и умноженный на 100%.

Для того чтобы все значения показателей принадлежали области значения активационной функции, а именно, гиперболического тангенса, до выполнения шагов алгоритма все множество наблюдений было нормировано при помощи сигмоидной функции:

$$\tilde{x}_i = f\left(\frac{x_i - \bar{x}}{\sigma}\right), f(\alpha) = \frac{2}{1 + e^{-\alpha}} - 1,$$

где x_i — значение показателя, \tilde{x}_i — нормированное значение, \bar{x} и σ — среднее значение показателя и его стандартное отклонение.

Далее, пользуясь вышеописанными методами, обучающее множество было сжато различными методами с различной степенью сжатия (от 20% до 100%, 20% означает, что в результате сжатия в множестве остается 20% наблюдений, 100% — что сжатие по сути не выполняется). На каждом из полученных множеств наблюдений были построены 2 нейросети:

1. 9 входов, соответствующих 9 определяющим показателям, 1 выход, 1 скрытый слой с 15 нейронами;
2. 10 входов, соответствующих 9 определяющим показателям и прогнозируемому показателю в текущем периоде, 1 выход, 1 скрытый слой с 15 нейронами.

На входах и скрытом слое в качестве активационной функции использовался гиперболический тангенс, для обучения нейросети применялся алгоритм упругого распространения.

В каждую построенную и обученную нейросеть были загружены все 616 наблюдений, и вычислены ошибки для каждого из наблюдений. Эти ошибки для каждого наблюдения были отождествлены с дополнительным условным показателем Z , который отображает общее воздействие (информационную суть прогноза) неучтенных в нейросети показателей качества среды обитания, влияющих на прогнозируемый показатель заболеваемости. Далее, были рассчитаны показатели z по каждой территории для прогнозируемого года, т.е. Z_{T+1} .

Далее для окончательного прогноза в каждую нейросеть были загружены показатели качества среды обитания и качества медицинского обслуживания по территориям за год, по которому необходимо было получить прогноз. После расчета нейросети к её прогнозному значению показателя заболеваемости были прибавлены значения Z_{T+1} для прогнозируемого временного периода по территориям. Эта сумма и являлась значением прогнозируемого показателя по территориям. Теперь осталось сравнить спрогнозированный и фактический показатели заболеваемости за 2014 год. Ошибки прогноза в процентах для каждого способа сжатия, степени сжатия и способа вычисления Z приведены в таблицах. В таблицах вариант «Без сжатия» и «Без Z » соответствует прогнозу с помощью нейросети без применения методов, описанных в данной статье.

Если значение прогнозируемого показателя не подается на вход нейросети, то лучший результат — 0.946%, который достигается при сжатии по прогнозируемому показателю до 20% и использовании последнего значения z . В случае использования значения прогнозируемого показателя в текущий период как входа нейросети лучший результат — 0.987%, который достигается при сжатии до 30% по прогнозируемому показателю и вычислении z с использованием сглаживания с параметром 3 и ΔZ . Т.е. ошибка прогнозирования благодаря данному методу уменьшается как минимум в 2 раза.

Из полученных результатов следует, что:

Сжатие не всегда повышает качество прогнозирования (сильное сжатие дает как самые лучшие, так и самые худшие результаты),

Вычисление z любым способом в конкретном примере повышает качество прогнозирования,

Сглаживание в большинстве случаев позволяет получить более точный результат,

В данном примере сильное сжатие по прогнозируемому показателю наиболее уменьшает ошибку прогноза,

Таблица 1. Результаты прогнозирования без учета значения прогнозируемого показателя в текущем периоде (приведены лучшие и худшие результаты, а также результат без применения метода).

Степень сжатия	Показатели, по которым выполняется сжатие	Без Z	Последний Z	Средний Z	Монотонность Z	Монотонность Z с ΔZ
20	Прогнозируемый	5,959	0,946	1,666	1,546	1,558
30	Прогнозируемый	5,863	0,967	1,651	1,476	1,477
40	Прогнозируемый	6,083	1,058	1,836	1,559	1,566
50	Прогнозируемый	5,482	0,998	1,658	1,413	1,411
100	Без сжатия	4,043	1,766	2,912	2,718	2,723
40	Определяющие	7,951	2,405	3,925	3,415	3,399
20	Определяющие	12,685	2,746	4,346	3,975	3,961
30	Определяющие	17,782	2,945	4,54	4,472	4,472

Продолжение таблицы

Степень сжатия	Показатели, по которым выполняется сжатие	Сглаживание Z при k=2	Сглаживание Z при k=3	Сглаживание Z при k=2 с ΔZ	Сглаживание Z при k=3 с ΔZ
20	Прогнозируемый	1,189	1,217	1,186	1,216
30	Прогнозируемый	1,071	1,067	1,069	1,063
40	Прогнозируемый	1,206	1,228	1,201	1,222
50	Прогнозируемый	1,2	1,263	1,193	1,257
100	Без сжатия	2,168	2,327	2,159	2,329
40	Определяющие	2,812	2,997	2,78	2,969
20	Определяющие	3,157	3,261	3,157	3,267
30	Определяющие	3,847	3,752	3,844	3,757

Таблица 2. Результаты прогнозирования с учетом значения прогнозируемого показателя в текущем периоде (приведены лучшие и худшие результаты, а также результат без применения метода).

Степень сжатия	Показатели, по которым выполняется сжатие	Без Z	Последний Z	Средний Z	Монотонность Z	Монотонность Z с ΔZ
30	Прогнозируемый	4,458	1,014	1,227	1,221	1,228
40	Прогнозируемый	3,457	1,031	1,212	1,176	1,176
90	Прогнозируемый	3,025	1,065	1,122	1,144	1,147
50	Прогнозируемый	3,54	1,052	1,195	1,164	1,166
100	Без сжатия	1,883	1,238	1,16	1,181	1,185
20	Определяющие	6,657	1,393	1,736	1,674	1,673
30	Определяющие	9,256	1,468	1,922	1,735	1,747

Продолжение таблицы

Степень сжатия	Показатели, по которым выполняется сжатие	Сглаживание Z при k=2	Сглаживание Z при k=3	Сглаживание Z при k=2 с ΔZ	Сглаживание Z при k=3 с ΔZ
30	Прогнозируемый	1,003	0,99	1,002	0,987
40	Прогнозируемый	1,039	1,037	1,036	1,034
90	Прогнозируемый	1,048	1,033	1,047	1,032
50	Прогнозируемый	1,072	1,038	1,071	1,035
100	Без сжатия	1,248	1,24	1,245	1,237
20	Определяющие	1,483	1,499	1,484	1,496
30	Определяющие	1,435	1,457	1,437	1,455

Метод позволяет получить примерно одинаковую точность прогнозирования, как при использовании, так и без использования текущего значения прогнозируемого показателя как входа нейросети.

Заключение

Применение метода сжатия-расширения информационного пространства для прогноза какого-либо показателя с помощью экстраполяции ошибки нейросети на практике является очень трудоёмким с большими временными затратами. Однако в некоторых «безнадёжных» случаях даёт весьма удовлетворительные результаты прогноза. Поэтому авторы разработали комплекс программ, позволяющих существенно сократить временные затраты на построение оптимальной прогнозной нейросетевой модели. Программный комплекс позволяет сравнивать различные сценарии сжатия при всех остальных равных условиях (степень сжатия, архитектура нейронной сети, способ экстраполирования ошибки). Сравнить степени сжатия для каждого сценария сжатия при всех остальных равных условиях (архитектура нейронной сети, способ экстраполирова-

ния ошибки). Сравнить способы экстраполирования ошибки при всех остальных равных условиях (сценарий сжатия, степень сжатия, архитектура нейронной сети). Таким образом, используя программный комплекс можно легко установить для любого прогнозируемого показателя точный сценарий сжатия и точный способ экстраполяции ошибки нейронной сети. Также можно установить примерную степень сжатия информационного пространства. В программном комплексе анализируются следующие степени сжатия: 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% (то есть без сжатия). И примерную архитектуру нейронной сети (количество скрытых слоёв и количество нейронов для каждого скрытого слоя, активационные функции, алгоритм обучения и количество итераций обучения). После чего построение оптимальной для прогноза нейросети становится для исследователя делом опыта нейросетевого моделирования.

В дальнейшем авторам кажется, очень важным установить робастность или устойчивость алгоритма построения нейронной сети для прогноза того или иного показателя. Об этом будет следующая публикация.

ЛИТЕРАТУРА

1. Ruta D., Gabrys B. Neural Network Ensembles for Time Series Prediction. Proc. Int'l Joint Conf. Neural Networks. 2007, P. 1204–1209.
2. Dolenko S., Guzhva A., Persiantsev I., Shugai J. Multi-stage Algorithm Based on Neural Network Committee for Prediction and Search for Precursors in Multidimensional Time Series. In: Alippi C., Polycarpou M., Panayiotou C., Ellinas G. Artificial Neural Networks — ICANN2009. Lecture Notes in Computer Science. 2009, v. 5769, P. 295–304.
3. Гусев А. Л. Управление и процедуры групповых проверок. Saarbrücken (Germany): LAP LAMBERT Academic Publishing GmbH & Co. KG, 2011. — 75 с.
4. Гусев А. Л. Непрерывный статистический контроль при управлении. Los Angeles (USA): Createspace, 2012. — 128 с.
5. Kalaivani R, Sudhagar K, Lakshmi P. Neural Network based Vibration Control for Vehicle Active Suspension System. Indian Journal of Science and Technology. 2016, Jan; 9(1).
6. Гусев А. Л. Нейронные сети и прогноз. Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века: сб. ст. по материалам Всероссийской научно-практической конференции (г. Пермь, 17–19 мая 2016 г.) / ПГНИУ. — Пермь, 2016. — С. 60–62.
7. Collotta M., Pau G. An Innovative Approach for Forecasting of Energy Requirements to Improve a Smart Home Management System Based on BLE. IEEE Transactions on Green Communications and Networking. 2017, v. 1, № 1, P. 112–120.
8. Raji C. G., Vinod Chandra S. S. Long-Term Forecasting the Survival in Liver Transplantation Using Multilayer Perceptron Networks. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2017, v. 47, № 8, P. 2318–2329.
9. Gusev A. L., Okunev A. A. Forecasting with incomplete set of factors determining the predicted factor. Neural network error extrapolation method. International Journal of Applied Mathematics and Statistics. 2017, v. 56, № 5, P.48–52.
10. Регионы России. Социально-экономические показатели 2015. Статистический сборник: Росстат. — М., 2015. — 1266 с.

© Гусев Андрей Леонидович (alguseval@mail.ru), Окунев Александр Анатольевич (alexander2510@mail.ru).

Журнал «Современная наука: актуальные проблемы теории и практики»