

УДК 007:159.955:519.72

DOI:10.25729/ESI.2024.35.3.003

Пластичность самоподобных нейронных сетей

Дорогов Александр Юрьевич

Санкт-Петербургский государственный электротехнический университет,

ПАО «Информационные телекоммуникационные технологии»,

Россия, Санкт-Петербург, vaksa2006@yandex.ru

Аннотация. В работе исследуется пластичность многослойных модульных нейронных сетей, обладающих характеристическим свойством самоподобия. Для оценки пластичности используется известное из механики понятие степеней свободы. Число степеней свободы сети оценивается максимальной размерностью операторного многообразия нейронной сети, образованного вариацией параметров нейронных модулей и наличием межмодульных связей. Для получения оценок пластичности нейронные модули рассматриваются как линейные операторы фиксированного ранга. Получены расчетные формулы для вычисления размерности операторного многообразия нейронного модуля вне и в составе сети. Нейронная сеть рассматривается, как двойственный оператор сложной структуры, входом и выходом которого являются векторные пространства. На уровне структурной модели вводится понятие модальных состояний сети, характеризующих размерности векторных подпространств на входе и выходе нейронных модулей в составе сети. Оценка размерности многообразия сети выполняется через ее модальные состояния. Отмечено, что самоподобные сети принадлежат классу слабосвязанных сетей, для которых расчет модальных состояний не вызывает трудностей. Получены точные формулы расчета степени пластичности слабосвязанных нейронных сетей. Результаты анализа использованы для оценки пластичности быстрых нейронных сетей (БНС), и их подмножества - пирамидальных БНС прямой и обратной ориентации.

Ключевые слова: Нейронная сеть, структурная модель, самоподобие, модальные состояния, пластичность, степени свободы

Цитирование: Дорогов А.Ю. Пластичность самоподобных нейронных сетей / А.Ю. Дорогов // Информационные и математические технологии в науке и управлении, 2024. – № 3(35). – С. 33-43. – DOI: 10.25729/ESI.2024.35.3.003.

Введение. В нейрофизиологии под пластичностью биологических нейронных сетей понимают способность нейронов сети изменять свои свойства под воздействием внешней среды. В применении к искусственным нейронным сетям это понятие можно трактовать как способность нейронной сети к обучению. В работе [1] автором впервые была предложена оценка пластичности для многослойных нейронных сетей прямого распространения. В основу оценки было положено известное из механики понятие числа степеней свободы.

При анализе пластичности многослойная нейронная сеть рассматривается как некоторая точка-оператор в пространстве операторов. Изменение синаптических весов сети приводит к перемещению этой точки в операторном пространстве. Число возможных степеней свободы операторной точки может служить характеристикой пластичности сети. Данная оценка имеет локальный характер, глобальной оценкой может служить верхняя граница числа степеней свободы по всем допустимым точкам операторного пространства.

Следует отметить, что задача оценки степеней свободы перестраиваемых операторов давно вызывала интерес исследователей, еще в 1971 году Andrews H.C. и Caspari K.L. [2] предложили метод расчета числа степеней свободы для линейных операторов, полученных произведениями слабозаполненных квадратных матриц. Суть метода сводилась к прямому подсчету всех варьируемых коэффициентов, однако, при этом не учитывалось, что при перемножении матриц варьируемые коэффициенты становятся зависимыми, в результате исследователи получили завышенный результат. С позиций сегодняшнего дня матричные произведения, использованные Andrews H.C. и Caspari K.L., можно трактовать, как многослойные нейронные сети с линейными функциями активации и ограниченными связями.

Для нейронных сетей с произвольной структурой в настоящее время неизвестен способ вычисления оценки пластичности, однако для самоподобных нейронных сетей прямого распространения с ядерной организацией [3, 4] такую оценку можно получить для ряда практически важных случаев. Число степеней свободы нейронной сети можно интерпретировать как размерность её операторного многообразия, порождаемого вариацией всех возможных параметров сети. При этом предполагается, что операторное многообразие нейронной сети локально дифференцируемо по всем координатам, т. е. является непрерывным и гладким. Гладкое многообразие локально устроено, как линейное пространство [5], и его размерность оценивается через максимальную размерность локальных пространств. Понятно, что данная оценка не изменится, если вместо полного операторного многообразия нейронной сети рассматривать только его часть (трансверсальную область), в которой размерность локального пространства достигает максимума. Эту выделенную часть обычно называют областью «общего положения».

Концепция «общего положения» в моделировании является выражением фундаментального понятия «типичности» при совместном размещении геометрических образов в векторном пространстве. Операторное многообразие нейронной сети формируется взаимным объединением операторных многообразий нейронных модулей. За счет существующих межмодульных связей эти многообразия имеют множество пересечений. Для оценки пластичности необходимо локализовать трансверсальную область результирующего многообразия и определить ее размерность. Локально каждое операторное многообразие устроено, как некоторое линейное подпространство в пространстве операторов, поэтому многообразия локально пересекаются по векторным подпространствам. Для двух подпространств L_1, L_2 векторного пространства L под общим положением понимается [6] такое их размещение, при котором их пересечение имеет минимальную, а сумма – максимальную размерность (под суммой или объединением подпространств понимается их линейная оболочка). Другой термин для этого же понятия – L_1, L_2 пересекаются трансверсально. Подпространства в «общем положении» либо не пересекаются, а если пересекаются, то размерность их объединения равна:

$$\dim L_1 \cup L_2 = \dim L_1 + \dim L_2 - \dim L_1 \cap L_2. \quad (1)$$

Название «общее положение» обусловлено тем, что в некотором смысле большинство пар подпространств L_1, L_2 находятся в общем положении, а другие расположения являются вырожденными.

Поскольку размерность многообразия оценивается через локальную размерность линейного пространства, то в контексте оценки пластичности можно ограничиться линейными операторами. Для модульной нейронной сети задача расчёта степени пластичности разделяется на две подзадачи: в первой требуется определить пластичность отдельных модулей, а во второй – используя полученные данные и информацию о структуре сети, определить пластичность всей сети, следуя при этом принципу трансверсальности при объединении операторных многообразий. Рассмотрим обе эти задачи последовательно.

2. Операторные многообразия нейронных модулей. Будем полагать, что в контексте оценки пластичности нейронной сети нейронный модуль описывается линейным оператором. Пусть E и D – ассоциированные пространства входа и выхода для модуля A с размерностями p и g по входу и выходу. Ограничиваясь моделью общего положения, будем считать, что при варьировании параметров операторное многообразие модуля совпадает с многообразием линейных отображений $A_r: E \rightarrow D$ ранга r .

Как известно [7], полное множество линейных операторов, действующих из пространства E в D , изоморфно тензорному произведению $E \otimes D$, которое является линейным пространством размерности $p \cdot g$. Согласно теореме о структуре линейного отображения [6] для каждого оператора ранга r существуют такие прямые разложения пространств:

$$E = E_0 \oplus E_1 \Leftrightarrow D_1 \oplus D_0 = D, \tag{2}$$

что D_1 изоморфно E_1 и имеет размерность r , а E_0 составляет ядро отображения. Подпространство D_0 называют коядром, оно характеризует степень неопределённости (неоднозначности) оператора.

Из разложения (2) нетрудно видеть, что подмножество операторов A_r представляет собой объединение подмножества операторов ранга r , осуществляющих отображение из пространства E в D_1 , и подмножества операторов ранга r , осуществляющих отображение из E_1 в D . На языке тензорных произведений это можно записать так:

$$A_r \cong (E \otimes D_1) \cup (E_1 \otimes D).$$

Отсюда, следуя правилу вычисления размерности объединения подпространств общего положения (1), получим:

$$\dim A_r = \dim(E \otimes D_1) + \dim(E_1 \otimes D) - \dim(E \otimes D_1) \cap (E_1 \otimes D). \tag{3}$$

Пересечением множеств является подмножество операторов, осуществляющих отображение из подпространства E_1 в D_1 , поэтому

$$\dim(E \otimes D_1) \cap (E_1 \otimes D) = r^2.$$

Поскольку $\dim E = p$, $\dim D = g$, $\dim E_1 = \dim D_1 = r$, то из (3) следует:

$$\dim A_r = pr + gr - r^2. \tag{4}$$

В частности, для операторного многообразия полного ранга, когда $r = \min(p, g)$, из данного выражения получим очевидное значение $\dim A_r = p \cdot g$. Аналогичные соотношения для определения размерности многообразия матриц ранга r , представлены в работе [8].

3. Модуль двойственного функционирования в составе сети. Обозначим $A_r^\# : D \rightarrow E$ множество всех операторов обратного распространения ранга r , действующих из пространства D в пространство E . Разложение (2) симметрично для класса прямых и обратных отображений, поэтому также представляет собой многообразие и существует естественный изоморфизм $A_r \cong A_r^\#$, который задаётся совпадением пар (E_1, D_1) . Данный изоморфизм является выражением двойственности в представлении модуля (см. рис. 1).

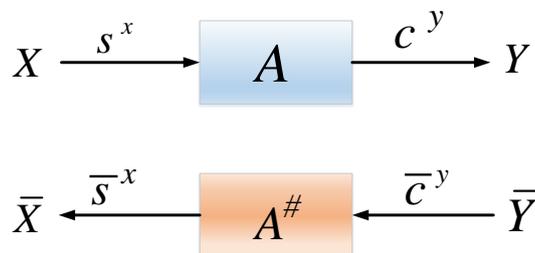


Рис. 1. Нейронные модули в прямой и двойственной сети

Возможность тензорного представления множества операторов позволяет в расчетах перейти от операторного многообразия к одному оператору фиксированного ранга, отображающего входное пространство векторов в выходное векторное пространство. Зная размерности этих пространств для прямого и двойственного операторов, размерность многообразия операторов можно вычислить, следуя выражению (4). Условия общего положения теперь предъявляются к корректному выбору оператора.

Таким образом, будем полагать, что операторный модуль общего положения ранга r , размерностью p и g по входу и выходу находится под воздействием пространств E и D . В режимах прямого и обратного функционирования размерность выходного пространства операторного модуля общего положения не может быть больше r . Следовательно, модуль может реализовать не более, чем pr степеней свободы при прямом функционировании и не более, чем gr – при обратном. При этом r^2 степеней свободы являются общими. Дуальный переход от многообразия операторов к векторным пространствам, ассоциированным с каждым модулем сети, естественным образом вводит сигнальные состояния на структурной модели нейронной сети. Размерность сигнальных подпространств в различных сечениях двойственной сети теперь будем рассматривать как модальные состояния структурного уровня (от латинского *modus* – размер). Операторные модули общего положения осуществляют преобразование входного модального состояния в выходное в прямой и двойственной сетях.

Если модуль находится в составе сети, то размерности входных сигнальных пространств $X \subseteq E$ и $\bar{Y} \subseteq D$ в прямой и двойственной сетях в общем случае меньше максимально возможных (т.е. p и g). Это ограничивает возможности по реализации степеней свободы модуля. Действующее число степеней свободы модуля в составе сети по аналогии с (4) будет равно

$$\dim \hat{A}_r = s^x \hat{r} + \bar{c}^y \hat{r} - \hat{r}^2, \quad (5)$$

где s^x, \bar{c}^y – размерности входных подпространств в прямой и двойственной сетях, $\hat{r} = \min(r, s^x, c^y)$ – действующий ранг нейронного модуля.

При системном анализе наибольший интерес представляют потенциально достижимые возможности нейронной сети. Понятно, что с увеличением операторных рангов модулей такие возможности увеличиваются, поэтому при анализе можно полагать, что всегда выполняется равенство $r = \min(p, g)$. В дифференциальной геометрии такой оператор называется регулярным или правильным; этот же термин будем использовать для нейронного модуля. Для регулярного модуля в составе сети выполняется условие $r \geq \min(s^x, c^y)$ и, следовательно, $\hat{r} = \min(s^x, \bar{c}^y)$, тогда из (5) нетрудно получить, что

$$\dim \hat{A}_r = s^x \bar{c}^y. \quad (6)$$

Для операции нахождения минимума в дальнейшем будем использовать обозначение « \circ », иначе называемое логическим умножением. Например, с этим обозначением условие регулярности модуля будет записываться в виде $r = p \circ g$.

4. Пластичность модульных нейронных сетей с биективными связями.

Будем полагать, что многообразие операторов в модульной нейронной сети образуется только за счёт вариации параметров модулей. Межмодульные связи при этом считаются фиксированными.

4.1. Влияние модулей. Выделим в сети некоторый модуль A ранга r и предположим, что его параметры варьируются, в то время как параметры всех остальных модулей фиксированы в общем положении. Поскольку модуль находится в составе сети, то размерности пространств состояний модуля в прямой и двойственной сетях в общем случае меньше размерностей модуля по входу и выходу. Действующее число степеней свободы модуля обозначим $S(A)$. Выделяя подобным образом и поодиночно варьируя параметры остальных модулей, получим, что вклад всех модулей в общее число степеней свободы сети определяется суммой:

$$\sum_A S(A).$$

4.2. Влияние связей. Биективные межмодульные связи устанавливают точное и однозначное соответствие между модальными состояниями смежных модулей, т. е. каждая связь, рассматриваемая как двойственное отображение, является регулярной и её влияние можно оценивать выражением (6). Однако, в отличие от модулей, связи фиксированы, и поэтому их присутствие в сети приводит к «связыванию» степеней свободы смежных нейронных модулей. Из этого следует, что каждая фиксированная межмодульная связь (допустим, между модулями $A \Rightarrow B$) ранга r^{AB} уменьшает общее количество степеней свободы сети на величину

$$s^B \bar{c}^A = (c^A \circ r^{AB})(\bar{s}^B \circ r^{AB}),$$

где s^B, c^A и \bar{c}^A, \bar{s}^B размерности пространств модальных состояний в прямой и двойственной сетях (вторая форма записи более удобна при конусе связей). Рис. 2 иллюстрирует действие фиксированной межмодульной связи в сети двойственного распространения.

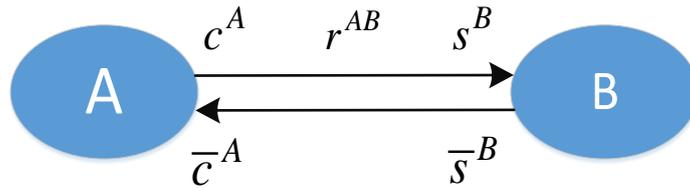


Рис. 2. Межмодульная связь в нейронной сети двойственного распространения

Таким образом, формула расчёта числа степеней свободы для всей модульной сети будет иметь вид:

$$S(H) = \sum_A S(A) - \sum_{A \Rightarrow B} \sum (c^A \circ r^{AB})(\bar{s}^B \circ r^{AB}), \tag{7}$$

а если все модули регулярны, то:

$$S(H) = \sum_A s^A \bar{c}^A - \sum_{A \Rightarrow B} \sum (c^A \circ r^{AB})(\bar{s}^B \circ r^{AB}). \tag{8}$$

В последних выражениях двойная сумма распространяется на все существующие связи.

5. Расчёт модальных состояний в самоподобных нейронных сетях.

Выражения (7), (8) можно использовать для расчёта степени пластичности только в том случае, если известны модальные состояния нейронной сети. Модальные состояния в прямой и двойственной сетях порождаются наличием на их входах сигнальных пространств. Для оценки потенциальных свойств нейронной сети будем полагать, что эти пространства имеют максимально возможную размерность, определяемую размерностью терминальных полей сети.

$$s^{inp} = N, \quad \bar{c}^{out} = M.$$

Здесь N – размерность нейронной сети по входу, M – размерность нейронной сети по выходу. Напомним, что выход прямой сети является одновременно входом двойственной сети, а состояния s^{inp}, \bar{c}^{out} являются источниками, воздействующими на прямую и двойственную сеть. Схема формирования модальных состояний двойственной сети показана на рис. 3.

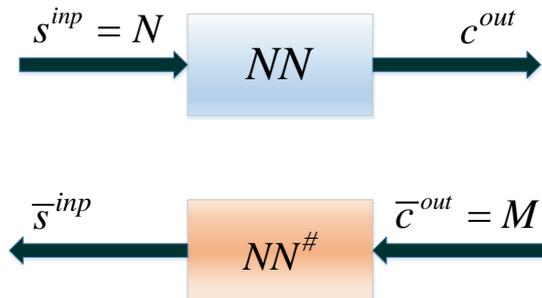


Рис. 3. Схема формирования модальных состояний в прямой и двойственной нейронной сети

Рассмотрим конус связей к некоторой вершине B в модульной сети прямого распространения (см. рис. 4). Пусть эта вершина соответствует оператору ранга r^B с размерностью p^B по входу и g^B по выходу (рис. 5).

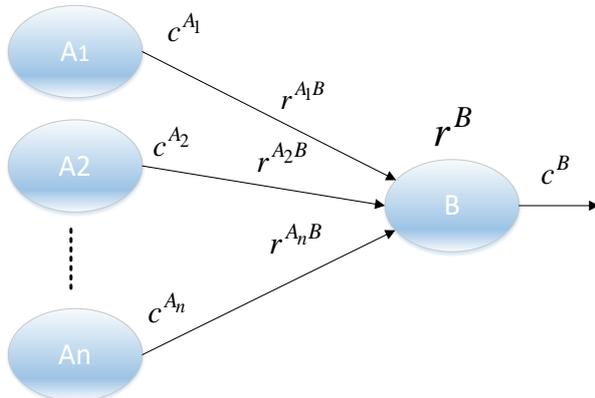


Рис. 4. Конус связей в сети прямого распространения

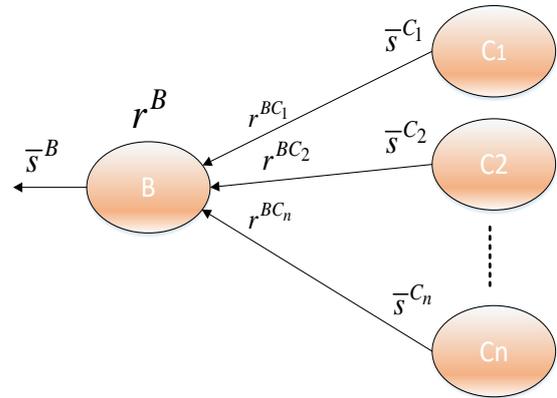


Рис. 5. Конус связей в сети обратного распространения

Модальные состояния для конуса связей в общем виде определяются выражениями:

$$c^B = s^B \circ r^B, \quad s^B = \bigcup_{A_i \in \Gamma^{-1}(B)} c^{A_i} \circ r^{A_i B},$$

где символ \bigcup условно показывает, что размерность s^B следует определять для объединения подпространств, порождаемых модулями окрестности $\Gamma^{-1}(B)$. Размерности этих подпространств известны и равны $c^{A_i} \circ r^{A_i B}$, однако размерность их объединения в общем случае вычислить не удаётся, поскольку в произвольной структуре пространства могут иметь пересечения, обусловленные наличием предшествующих связей между модулями.

В случае самоподобных сетей задача существенно упрощается, поскольку по теореме «О параллельных путях» [3] для любого модуля самоподобной сети существует единственный путь, связывающий его с терминальной вершиной. Этот путь выделяет в терминальном пространстве независимое подпространство. Размерность этого подпространства легко вычислить: оно будет равно логическому произведению рангов всех дуг пути от терминальной вершины до данного модуля. Поскольку все выделенные таким образом подпространства не пересекаются, то объединение подпространств перед модулем B фактически будет прямой суммой, а это означает, что формула для вычисления размерности трансформируется к виду:

$$c^B = s^B \circ r^B, \quad s^B = \sum_{A_i \in \Gamma^{-1}(B)} c^{A_i} \circ r^{A_i B} = p^B. \tag{9}$$

Поскольку модули регулярны, то $c^{A_i} \circ r^{A_i B} = r^{A_i B}$. Более того, в самоподобных сетях все связи инъективны, поэтому сумма (9) будет равняться входной размерности модуля, т.е. $s^B = p^B$. Следуя принципу двойственности, можно записать аналогичные выражения для расчёта модальностей для сети обратного распространения (см. рис.5):

$$\bar{s}^B = \bar{c}^B \circ r^B, \quad \bar{c}^B = \sum_{C_i \in \Gamma(B)} \bar{s}^{C_i} \circ r^{C_i B} = g^B, \tag{10}$$

где g^B - размерность модуля по выходу.

5.1. Пластичность быстрых нейронных сетей. Быстрое перестраиваемое преобразование [3] будем рассматривать как модульную быструю нейронную сеть (БНС), где роль модулей выполняют базовые операции (нейронные ядра). Эта сеть является самоподобной. По построению быстрого преобразования ранги всех межмодульных связей равны единице, а все ядра в пределах слоя имеют совпадающие структурные характеристики. Обозначим через

(p_m, g_m) размерности ядер слоя m , а через k_m – их число. Нетрудно заметить, что при единичных рангах каждая фиксированная связь удаляет одну степень свободы. Учитывая это обстоятельство и используя (9), (10) из формулы (8), получим:

$$S(H) = \sum_{m=0}^{n-1} p_m g_m k_m - \sum_{m=0}^{n-2} D_m, \quad (11)$$

где D_m – количество связей в межслойном переходе с номером m , n – число нейронных слоев. Например, для сети, топологическая и структурная модели которой показаны на рис. 6 и рис.7, число степеней свободы равно $S(H) = (3 \cdot 2 \cdot 4 + 2 \cdot 2 \cdot 4 + 2 \cdot 2 \cdot 4) - (8 + 8) = 40$.

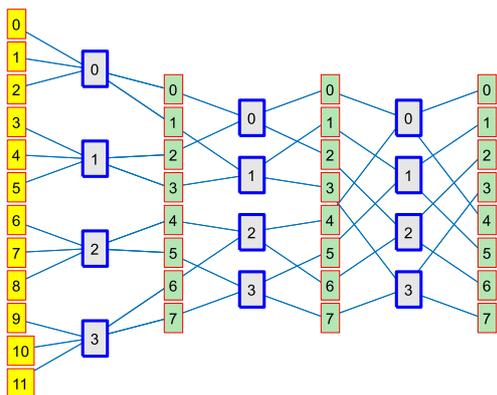


Рис. 6. Топологическая модель трехслойной БНС

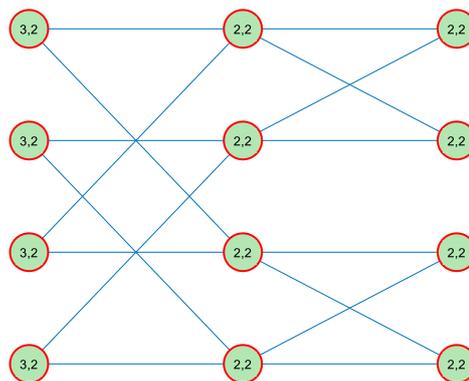


Рис. 7. Структурная модель трехслойной БНС

5.2. Пластичность самоподобной пирамидальной нейронной сети. Ключевая идея самоподобных пирамидальных нейронных сетей [9] заключается в изменении топологии БНС таким образом, чтобы выделить сетевые компоненты, информационная ёмкость которых используется полностью и отсечь компоненты, информационная ёмкость которых не используется при обучении, сохранив при этом условия самоподобия. В результате нейронная сеть упрощается, быстродействие её увеличивается, и глубина обучения становится максимально-возможной. На рис. 8 и рис. 9 показаны топологическая и структурная модели самоподобной пирамидальной нейронной сети. На структурной модели ранги всех связей равны 1.

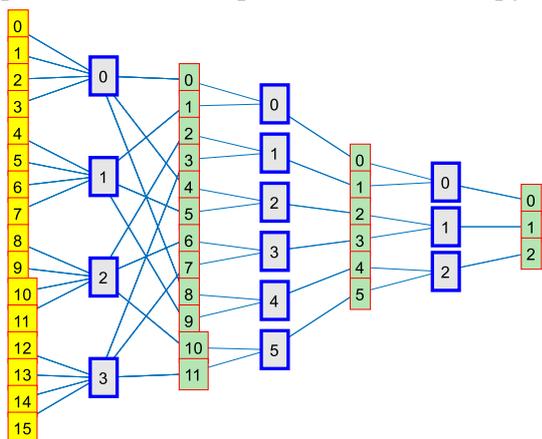


Рис. 8. Топологическая модель пирамидальной самоподобной сети

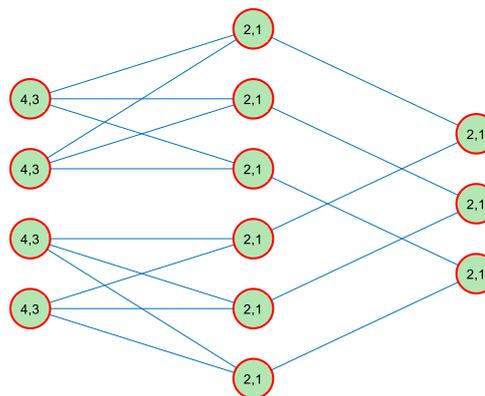


Рис. 9. Структурная модель пирамидальной самоподобной сети

Данная сеть может обучиться к трём эталонным функциям, заданным на интервале длиной 16. Используя расчетную формулу (11), непосредственно из структурной модели получим

$S(H) = 4 \cdot 3 \cdot 4 + 2 \cdot 1 \cdot 6 + 2 \cdot 1 \cdot 3 - 12 - 6 = 48$. При произвольном выборе трёх функций необходимо задать 48 значений, что совпадает с полученным числом степеней свободы пирамидальной сети. Таким образом, рассматриваемая сеть является глубокой в том смысле, что её потенциал обучения используется полностью, и покрывает все допустимое многообразие эталонных сигналов.

5.3. Информационная емкость нейросетевой памяти образов. В [9] показано, что самоподобная пирамидальная сеть обратной ориентации может быть эффективно использована для хранения одномерных и многомерных образов. Пластичность нейронной сети здесь трактуется как информационная емкость памяти. На рис. 10 и рис. 11 приведен пример нейросетевой памяти в топологическом и структурном представлении пирамидальной сети. Ранги всех связей структурной модели равны 1.

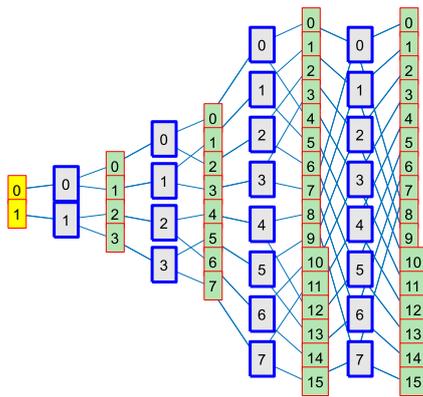


Рис. 10. Топологическая модель нейросетевой памяти образов

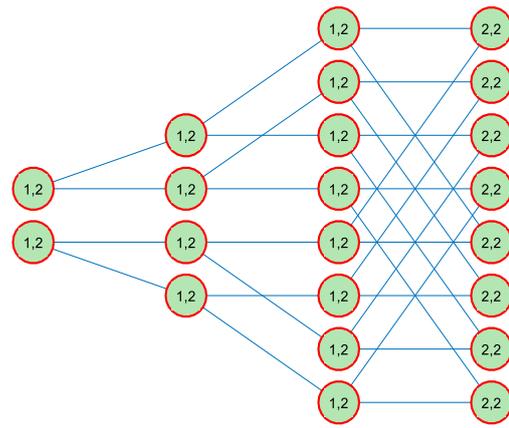


Рис. 11. Структурная модель нейросетевой памяти образов

Данная нейронная сеть обеспечивает хранение двух произвольных дискретных функций, заданных на интервале длиной $M = 16$. Используя расчетную формулу (11), непосредственно из структурной модели получим $S(H) = (1 \cdot 2) \cdot 14 + (2 \cdot 2) \cdot 8 - 28 = 32$. Для задания двух функций на интервале длиной 16 требуется определить 32 значения. Таким образом, сеть полностью использует свой потенциал для хранения функций и поэтому относится к категории сетей с глубокой степенью обучения. В отличие от кристаллов цифровой памяти с последовательным хранением данных, где восстановление образа происходит за счёт последовательного опроса ячеек, в памяти на БНС все пиксели образа восстанавливаются одновременно, что потенциально обеспечивает сверхвысокое быстродействие при считывании памяти.

Заключение. Пластичность нейронных сетей характеризует дифференциальную способность нейронной сети к обучению. Можно сказать, что качество одной нейронной сети лучше другой, если первая при прочих равных условиях обладает лучшей пластичностью. Но важно заметить, что дифференциальные оценки дают только линейное приближение для локальной области операторного пространства нейронной сети. Нелинейные свойства нейронной сети при этом игнорируются. Можно считать, что полученные численные значения являются нижней границей для оценки информационной емкости сети. В настоящее время не существует методов, позволяющих корректно учесть вклад нелинейностей в интеллектуальный потенциал нейронных сетей произвольной архитектуры. Однако наблюдается тенденция упрощения нелинейных элементов, например, замена их полулинейными элементами типа ReLU [10], что во многом обусловлено технологическими ограничениями при аппаратной реализации нейронных сетей. Элементы ReLU в этом случае выполняют роль управляемых ключей, а нейронная сеть работает в кусочно-линейном режиме.

В данной работе получены формулы расчета степени пластичности самоподобных многослойных нейронных сетей. Приведенные выражения позволяют определить точное значение степени пластичности на основе структурных характеристик нейронной сети. Расчетные формулы базируются на теореме «О параллельных путях», которая утверждает, что структурные модели самоподобных сетей не содержат параллельных путей. Сети, в которых отсутствуют параллельные пути, называются также слабосвязанными [11, 12]. Класс слабосвязанных сетей несколько шире класса самоподобных сетей, в частности, к этому классу относятся полносвязанные многослойные сети, однослойные персептроны, двухслойные ядерные сети с произвольной структурой. Полученные расчетные формулы применимы ко всему классу регулярных слабосвязанных сетей. В [11] представлен также графический расчет числа степеней свободы для слабосвязанных сетей общего вида.

Список источников

1. Дорогов А.Ю. Пластичность многослойных нейронных сетей / А.Ю. Дорогов // Известия вузов. Приборостроение, 1998. – Т.41. – №4. – С.36-41.
2. Andrews H.C., Caspari K.L. Degrees of freedom and modular structure in matrix multiplication. IEEE. Tr. Compt, 1971, vol. C-20, pp.113-141.
3. Дорогов А.Ю. Быстрые преобразования и самоподобные нейронные сети глубокого обучения. Часть 1. Стратифицированные модели самоподобных нейронных сетей и быстрых преобразований / А.Ю. Дорогов // Информационные и математические технологии в науке и управлении, 2023. – № 4(32). – С. 5-20. – DOI: 10.25729/ESI.2023.32.4.001.
4. Дорогов А.Ю. Быстрые преобразования и самоподобные нейронные сети глубокого обучения. Часть 2. Методы обучения быстрых нейронных сетей/ А.Ю. Дорогов // Информационные и математические технологии в науке и управлении, 2024.–№ 1(33). –С.5-19. –DOI:10.25729/ESI.2024.33.1.001.
5. Фоменко А.Т. Наглядная геометрия и топология: Математические образы в реальном мире – 2-е изд. / А.Т. Фоменко. – М.: Изд-во Моск.ун-та, Изд-во "ЧеРо", 1998. – 416 с.
6. Кострикин А.И. Линейная алгебра и геометрия: учеб. пособ. для вузов. 2-е изд. перераб. / А.И. Кострикин, Ю.И. Манин. – М.: Наука, 1986. – 304 с.
7. Ефимов Н.В. Линейная алгебра и многомерная геометрия / Н.В. Ефимов, Э.Р. Розендорн. – М.: Наука, 1970. – 528 с.
8. Дубровин Б.А. Современная геометрия: методы и приложения. Т. 2: Геометрия и топология многообразий. 5-е изд., испр. / Б.А. Дубровин, С.П. Новиков, А.Т. Фоменко. – М.: Эдиториал УРСС, 2001. – 296 с.
9. Дорогов А.Ю. Быстрые преобразования и самоподобные нейронные сети глубокого обучения. Часть 3. Пирамидальные нейронные сети с глубокой степенью обучения / А.Ю. Дорогов // Информационные и математические технологии в науке и управлении, 2024. – № 2(34). – С. 19-32. – DOI:10.25729/ESI.2024.34.2.002.
10. Liu Danqing A practical guide to relu. Medium, available at: <https://readmedium.com/a-practical-guide-to-relu-b83ca804f1f7> (accessed: 04/23/2024).
11. Дорогов А.Ю. Теория и проектирование быстрых перестраиваемых преобразований и слабосвязанных нейронных сетей. / А.Ю. Дорогов. – СПб.: Политехника, 2014. – 328 с.
12. Дорогов А. Ю. Самоподобные нейронные сети быстрого обучения/ А.Ю. Дорогов. – СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2024. – 188 с.

Дорогов Александр Юрьевич. Доктор технических наук, доцент, профессор кафедры «Автоматики и процессов управления» Санкт-Петербургского государственного электротехнического университета (СПбГЭТУ) «ЛЭТИ», главный научный сотрудник ПАО «Информационные телекоммуникационные технологии». Основные направления исследований автора связаны с интеллектуальным анализом данных, цифровой обработкой сигналов, проектированием быстрых преобразований и нейронных сетей быстрого обучения, разработкой аналитических платформ, моделированием радио-телекоммуникационных систем. AuthorID: 17611, SPIN: 8645-5873, ORCID: 0000-0002-7596-6761, vaksa2006@yandex.ru, г. Санкт-Петербург, ул. Попова, 5.

Plasticity of self-similar neural networks

Alexander Yu. Dorogov

St. Petersburg State Electrotechnical University,
PJSC “Information Telecommunication Technologies” (“Inteltech”),
Russia, St. Petersburg, *vaksa2006@yandex.ru*

Abstract. The plasticity of multilayer modular neural networks with the characteristic property of self-similarity is investigated in the paper. The concept of degrees of freedom, known from mechanics, is used to assess plasticity. The number of degrees of freedom of the network is estimated by the maximum dimension of the operator manifold of the neural network formed by variation of the parameters of neural modules and the presence of intermodule connections. To obtain plasticity estimates, neural modules are considered as linear operators of fixed rank. Calculation formulas for calculating the dimension of the operator manifold of a neural module outside and within the network are obtained. A neural network is considered as a dual operator of a complex structure, the input and output of which are vector spaces. At the level of the structural model, the concept of modal network states is introduced, characterizing the dimensions of vector subspaces at the input and output of neural modules in the network. The dimensionality of the network manifold is estimated through its modal states. It is noted that self-similar networks belong to a class of weakly coupled networks for which the calculation of modal states does not cause difficulties. Exact formulas for calculating the degree of plasticity of weakly coupled neural networks are obtained, the results of the analysis are used to assess the plasticity of fast neural networks (BNS), and their subsets - pyramidal BNS of direct and reverse orientation.

Keywords: Neural network, structural model, self-similarity, modal states, plasticity, degrees of freedom

References

1. Dorogov A.Yu. Plastichnost' mnogosloynnykh neyronnykh setey [Plasticity of multilayer neural networks]. *Izvestiya vuzov. Priborostroyeniye* [Journal of Instrument Engineering], 1998, vol 41, no. 4, pp. 36-41.
2. Andrews H.C., Caspari K.L. Degrees of freedom and modular structure in matrix multiplication. *IEEE. Tp. Compt*, 1971, vol. C-20, pp.113-141.
3. Dorogov A.Yu. Bystrye preobrazovaniya i samopodobnye neyronnye seti glubokogo obucheniya. Chast' 1. Stratifitsirovannye modeli samopodobnykh neyronnykh setey i bystrykh preobrazovaniy [Fast transformations and self-similar deep learning neural networks. Part 1. Stratified models of self-similar neural networks and fast transformations]. *Informatsionnyye i matematicheskiye tekhnologii v nauke i upravlenii* [Information and mathematical technologies in science and management], 2023, no. 4(32), pp.5-20, DOI:10.25729/ESI.2023.32.4.001.
4. Dorogov A.Yu. Bystrye preobrazovaniya i samopodobnye neyronnye seti glubokogo obucheniya. Chast' 2. Metody obucheniya bystrykh neyronnykh setey [Fast transformations and self-similar neural networks of deep learning. Part 2. Methods of training fast neural networks]. *Informatsionnyye i matematicheskiye tekhnologii v nauke i upravlenii* [Information and mathematical technologies in science and management], 2024, no. 1(33), pp. 5-19, DOI:10.25729/ESI.2024.33.1.001.
5. Fomenko A.T. Naglyadnaya geometriya i topologiya: Matematicheskie obrazy v real'nom mire. [Mathematical images in the real world]. 2nd ed. M., Izdatel'stvo Moskovskogo universiteta, Izdatel'stvo CheRo [Publ. Moscow university CheRo], 1998, 416 pp.
6. Kostrikin A.I., Manin Yu.I. Lineynaya algebra i geometriya: ucheb. posob. dlya vuzov [Linear algebra and geometry. Studies. help. for universities]. 2nd ed. Moscow, Nauka [Science], 1986, 304 pp.
7. Efimov N.V., Rozendorn E. R. Lineynaya algebra i mnogomernaya geometriya [Linear algebra and multidimensional geometry]. Moscow, Nauka [Science], 1970, 528 pp.
8. Dubrovin B.A., Novikov S.P., Fomenko A.T. Sovremennaya geometriya: metody i prilozheniya. T. 2: Geometriya i topologiya mnogoobraziy. [Modern geometry: methods and applications. Vol. 2: Geometry and topology of manifolds]. 5nd ed. Moscow, Editorial URSS, 2001, 296 pp.
9. Dorogov A.Yu. Bystrye preobrazovaniya i samopodobnye neyronnye seti glubokogo obucheniya Chast' 3. Piramidal'nye neyronnye seti s glubokoy stepen'yu obucheniya [Fast transformations and self-similar deep learning neural networks Part 3. Pyramid neural networks with a deep learning degree]. *Informatsionnyye i matematicheskiye tekhnologii v nauke i upravlenii* [Information and mathematical technologies in science and management], 2024, no. 2(34), no. 1(33), pp. 19-32, DOI:10.25729/ESI.2024.34.2.002.
10. Liu Danqing A practical guide to relu. Medium, available at: <https://readmedium.com/a-practical-guide-to-relu-b83ca804f1f7> (accessed: 04/23/2024).

11. Dorogov A.Yu. Teorija i proektirovanie bystryh perestraivaemyh preobrazovanij i slabosvjazannyh nejronnyh setej [Theory and design of fast tunable transformations and loosely coupled neural networks]. Spb. Publ Politehnika. 2014, 328pp.
12. Dorogov A.Yu. Samopodobnye nejronnye seti bystrogo obucheniya [Self-similar neural networks of fast learning]. SPb., Izd-vo SPbGETU «LETI» [St. Petersburg, Publishing house of SPbSETU “LETI”], 2024, 188 p.

Dorogov Alexander Yurievich. Doctor of technical sciences, associate professor, professor of the department of automation and control Processes of St. Petersburg state electrotechnical university (SPbETU) “LETI”, Chief researcher of PJSC “Information telecommunication technologies”. The main directions of the author's research are related to data mining, digital signal processing, the design of fast transformations and neural networks of fast learning, the development of analytical platforms, modeling of radio and telecommunications systems. AuthorID: 17611, SPIN: 8645-5873, ORCID: 0000-0002-7596-6761, vaksa2006@yandex.ru, St. Petersburg, st. Popova, 5.

Статья поступила в редакцию 12.05.2024; одобрена после рецензирования 23.07.2024; принята к публикации 30.09.2024.

The article was submitted 05/12/2024; approved after reviewing 07/23/2024; accepted for publication 09/30/2024.