

СТРУКТУРНЫЕ МОДЕЛИ И ТОПОЛОГИЧЕСКОЕ ПРОЕКТИРОВАНИЕ БЫСТРЫХ НЕЙРОННЫХ СЕТЕЙ.

Рассматриваются структурные и топологические модели многослойных быстрых нейронных сетей (БНС). Используется математический аппарат отображений числовых множеств и теории линейных представлений. Предложен алгоритм проектирования топологий БНС.

КЛЮЧЕВЫЕ СЛОВА: НЕЙРОННАЯ СЕТЬ, БЫСТРЫЙ АЛГОРИТМ, ТОПОЛОГИЯ СЕТИ, ЧАСТИЧНЫЕ ОТОБРАЖЕНИЯ, ПОДСТАНОВКИ, СТРУКТУРНАЯ МОДЕЛЬ.

Быстрые нейронные сети [1] являются разновидностью многослойных нейронных сетей прямого распространения. БНС сопоставимы с обычными нейронными сетями примерно в том же отношении как алгоритмы быстрого преобразования Фурье (БПФ) с прямым дискретным преобразованием Фурье. Высокая вычислительная эффективность БНС достигается за счет разумных ограничений на структурную организацию нейронной сети.

Структура БНС в какой-то мере повторяет структуру нейронных сетей живой природы, для которых всегда существуют ограничения на размерности рецепторных полей и на связи между нейронами. В работе [2] было показано что вопросы структурной организации БНС следует рассматривать на двух уровнях: уровне структурной модели БНС и уровне топологической реализации БНС. Если уровень структурной модели определяет общие свойства БНС по размерностям рецепторных полей и структуре межслойных связей, то уровень топологии определяет конкретную аппаратную или программную реализацию нейронной сети. Оба эти уровня рассмотрения неразрывно связаны между собой.

На рис.1 показан пример БНС в классическом представлении где каждая вершина соответствует одному нейрону, а дуги определяют связи между нейронами. Такое представление в дальнейшем будем называть топологической реализацией структурной модели. На рис.2 приведена структурная модель, построенная для данной сети. Каждой вершине структурной модели соответствует группа нейронов, имеющих общее рецепторное поле. Эти группы нейронов были названы нейронными ядрами. На структурном уровне каждое нейронное ядро характеризуется размерностью рецепторного поля и числом входящих в него нейронов. Эта пара чисел задает вес вершины структурной модели. Нейронное ядро j в слое m

описывается оператором A_{mj} , осуществляющего преобразования входного вектора собственного рецепторного поля:

$$Y_{mj} = (X_{mj})A_{mj}.$$

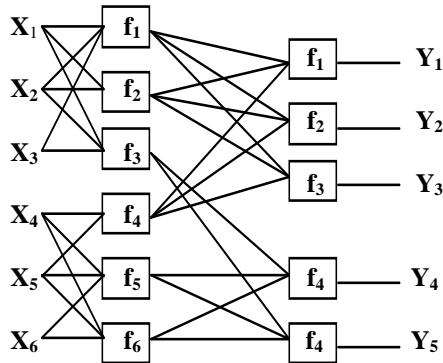


Рис. 1

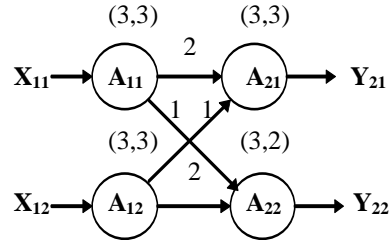


Рис. 2

В матричном представлении оператор нейронного ядра можно записать в виде:

$$S_{mj} = X_{mj}W_{mj}, \quad Y_{mj} = F(S_{mj}),$$

где W_{mj} - матрица синаптических весов ядра, $F()$ - многомерная функция активации, компонентами которой являются функции активации отдельных нейронов ядра.

Действие оператора нейронного ядра определено на паре градуированных [3] векторных пространств: пространстве рецепторов и пространстве нейронов слоя. Связи между слоями задаются проектирующими операторами, сохраняющими условия градуировки. Ранг проектирующего оператора определяет вес дуги структурной модели. Принципиальной особенностью построения БНС является независимость рецепторных полей ядер, т.е. рецепторные поля нейронных ядер не пересекаются.

Структурная модель является по существу описанием целого класса эквивалентных топологических реализаций БНС. Можно сказать, что множество топологий формирует орбиту структурной модели на множестве матричных представлений. Топология БНС задается числовыми частичными отображениями (частичными подстановками) [4]:

$$\sigma_j^m = \begin{pmatrix} u_1 & u_2 & \dots & u_{N_{xmj}} \\ 1 & 2 & \dots & N_{xmj} \end{pmatrix}, \quad \mu_j^m = \begin{pmatrix} v_1 & v_2 & \dots & v_{N_{ymj}} \\ 1 & 2 & \dots & N_{ymj} \end{pmatrix},$$

где N_{xmj} - размер рецепторного поля ядра j в слое m , N_{ymj} - число нейронов в ядре j слоя m . u - номер базисного вектора в пространстве рецепторов, v - номер базисного вектора в пространстве нейронов.

Действие частичных подстановок определено на входном и выходном векторах нейронного слоя, частичная подстановка σ_j^m выделяет рецепторное поля ядра, а подстановка μ_j^m нейронное поле. Формально действие частичных подстановок записывается в виде:

$$\begin{aligned} X_{mj} &= X_m * \sigma_j^m, \\ Y_{mj} &= Y_m * \mu_j^m. \end{aligned} \quad (1)$$

Комбинации символов $*\sigma_j^m$, и $*\mu_j^m$ можно рассматривать и как символическое обозначение проектирующих операторов, однако запись в виде формального произведения дает ряд преимуществ при математических выкладках. По принципу построения БНС проектирующие операторы не пересекаются, поэтому $\sigma_\alpha^m \wedge \sigma_j^m = 0$, $\mu_\alpha^m \wedge \mu_j^m = 0$ для любых $\alpha \neq j$.

Смежные нейронные слои связаны между собой перестановочными операторами перехода, действие которых определено подстановками q^m :

$$X_{m+1} = Y_m * q^m. \quad (2)$$

Операторы межслойного перехода индуцируют локальные операторы связи между нейронными ядрами, так что

$$X_{m+1,j} = \sum_i Y_{mi} * \rho_{ij}^m, \quad (3)$$

где ρ_{ij}^m - частичные подстановки соответствующие локальному оператору, \sum - символ прямой суммы векторов.

На уровне структурной модели каждая локальная связь характеризуется рангом проектирующего оператора. Числовое значение ранга определяет вес соответствующей дуги структурной модели (рис 2). Из выражений (1)-(3) следует:

$$\sigma_j^{m+1} = (q^m)^{-1} \sum_i \mu_i^m \rho_{ij}^m \quad (4)$$

Это выражение определяет рекуррентный алгоритм построения топологии нейронной сети.

Топологию нейронной сети компактно можно представить в виде произведения матриц смежности графа топологической реализации. Например, топологии нейронной сети, показанной на рис.1 отвечает произведение матриц вида:

$$\begin{vmatrix} 1 & 1 & 1 & & & \\ 1 & 1 & 1 & & & \\ 1 & 1 & 1 & & & \\ & & & 1 & 1 & 1 \\ & & & 1 & 1 & 1 \\ & & & 1 & 1 & 1 \end{vmatrix} \cdot \begin{vmatrix} 1 & 1 & 1 & & & \\ 1 & 1 & 1 & & & \\ & & & 1 & 1 & \\ 1 & 1 & 1 & & & \\ & & & & 1 & 1 \\ & & & & 1 & 1 \end{vmatrix} \quad (5)$$

Такие матрицы будем называть топологическими. Позиции ненулевых элементов в данных матрицах определяются частичными подстановками σ_j^m , μ_j^m . Нетрудно заметить, что при произвольной

перестановке элементов в любой строке частичных подстановок σ_j^m, μ_j^m топологические матрицы не изменяются. Неоднозначность в выборе частичных подстановок можно устранить если ограничиться множеством подстановок в которых отсутствуют инверсии [5]. при этом элементы строк будут всегда упорядочены по возрастанию.

При отсутствии инверсий отпадает необходимость в отображении вторых строк подстановок поскольку они всегда будут упорядочены по возрастанию. Таким образом расположение нейронного ядра в топологической матрице достаточно задать парой упорядоченных множеств U_j^m, V_j^m , образующих верхние строки частичных подстановок σ_j^m, μ_j^m . Введенные множества будем называть топологическими. Учитывая сказанное, построим удобный для программной реализации алгоритм построения топологии БНС. Приведем вначале описание алгоритма, а затем покажем, что он удовлетворяет общему соотношению (4).

Пусть для вершин A_{mj} топология задается парами (U_j^m, V_j^m) , а для вершин слоя $m+1$ парами (U_j^{m+1}, V_j^{m+1}) . Структурные связи между двумя смежными слоями запишем в виде ранговой матрицы:

$$R^m = \begin{vmatrix} r_{11}^m & r_{12}^m & \dots & r_{1l}^m \\ r_{21}^m & r_{22}^m & \dots & r_{2l}^m \\ \vdots & \vdots & \dots & \vdots \\ r_{k1}^m & r_{k2}^m & \dots & r_{kl}^m \end{vmatrix}$$

где r_{ij} - ранги проектирующих операторов межслойных связей. Каждой вершине слоя m соответствует строка ранговой матрицы, а каждой вершине слоя $m+1$ столбец матрицы. Сумма элементов строки i равна порядку множества V_i^m , а столбца j - порядку множества U_j^{m+1} . Введем в рассмотрение числовое множество $T^m = \{1, 2, \dots, N_m\}$, где $N_m = \text{card}(T^m) = \sum_i \sum_j r_{ij}^m$. (Нетрудно проверить, что значение N_m равно

числу нейронов в слое m и совпадает с размерностью рецепторного поля слоя $m+1$. Разместим элементы множества T^m в виде матрицы (рис 3.), подобной по структуре матрице R^m , причем размещение выполним так чтобы выполнялось условие $\text{card}(T_{ij}^m) = r_{ij}^m$. Подмножество элементов принадлежащих i -ой строке обозначим T_i^A , а j -ому столбцу T_j^B . Введем две произвольные подстановки на множестве T^m которые обозначим q^A и q^B . Тогда алгоритм построения топологических множеств будет определяться правилом:

$$V_i^m = T_i^A q^A, \quad U_j^{m+1} = T_j^B q^B$$

Графически алгоритм можно представить схемой, показанной на рис.3.

Покажем теперь, что данный алгоритм удовлетворяет соотношению (4). Из схемы рис. 3 следует:

$$\begin{array}{c}
 \left| \begin{array}{cccc} T_{11}^m & T_{12}^m & \dots & T_{1l}^m \\ T_{21}^m & T_{22}^m & \cdot & T_{2l}^m \\ \dots & \dots & \dots & \dots \\ T_{k1}^m & T_{k2}^m & \dots & T_{kl}^m \end{array} \right| \begin{array}{l} T_1^A \\ T_2^A \\ \dots \\ T_k^A \end{array} \xrightarrow{q^A} \begin{array}{l} V_1^m \\ V_2^m \\ \dots \\ V_k^m \end{array} \\
 \begin{array}{cccc} T_1^B & T_2^B & \dots & T_l^B \\ \downarrow q^B & \downarrow q^B & \dots & \downarrow q^B \\ U_1^{m+1} & U_2^{m+1} & \dots & U_l^{m+1} \end{array}
 \end{array}$$

Рис.3

$$U_j^{m+1} = \sum_i V_i^m q^m, \quad \text{где } q^m = (q^A)^{-1} q^B. \quad (6)$$

Поскольку $V_i^m = \{1, 2, \dots, N_{ymj}\} (\rho_{ij}^m)^{-1} (\mu_j^m)^{-1}$, а также

$$U_j^{m+1} = \{1, 2, \dots, N_{ymj}\} (\sigma_j^{m+1})^{-1}, \text{ то из (6) получим}$$

$$(\sigma_j^{m+1})^{-1} = \sum_i (\rho_{ij}^m)^{-1} (\mu_j^m)^{-1} q^m.$$

Обращение последнего выражения, очевидно, приводит к (4). Использование двух подстановок q^A и q^B вместо одной q^m делает алгоритм симметричным и более удобным при реализации.

Если $q^A = q^B$ то $q^m = (q^A)^{-1} q^B = e$, где e - тождественная подстановка.

Будем называть топологию для которой последнее равенство выполнено компактной и расширенной в противном случае. Для расширенной топологии между топологическими матрицами смежных слоев необходимо вводить перестановочную матрицу, соответствующую подстановке q^m . В общем случае для задания топологии необходимо фиксировать подстановки q^A , q^B и размещение множества T^m . При хранении топологии можно без потери информации сократить объем данных, если предварительно произвести «нормализацию» размещения множества T^m , следуя одному из ниже приведенных правил:

$$\begin{array}{ll}
 \text{a)} & T_{*a}^m = T^m q^A, \quad q_{*a}^A = e, \quad q_{*a}^B = (q^A)^{-1} q^B = q^m, \\
 \text{b)} & T_{*b}^m = T^m q^B, \quad q_{*b}^B = e, \quad q_{*b}^A = (q^B)^{-1} q^A = (q^m)^{-1}.
 \end{array} \quad (7)$$

Здесь символ «*» указывает значение после нормализации, а символы «a,b» варианты нормализации. Если топология компактная тогда нормализация любого типа приводит к равенству $q_{*}^A = q_{*}^B = e$. Таким образом для компактной топологии достаточно хранить только размещение

множества T^m , а для расширенной - размещение T^m и подстановку q^m . Покажем, что оба варианта нормализации эквивалентны, для этого выполним над нормализованным представлением по типу «*a» выполним нормализацию типа «*b»

$$T_{*a}^m q_{*a}^B = T^m q^A (q^A)^{-1} q^B = T^m q^B = T_{*b}^m,$$

$$(q_{*a}^B)^{-1} q_{*a}^A = (q_{*a}^B)^{-1} e = \left((q^A)^{-1} q^B \right)^{-1} = (q^B)^{-1} q^A = q_{*b}^A = (q^m)^{-1}.$$

Таким образом, мы приходим к нормализованному представлению типа «*b», определяемом правилом (7). Повторение нормализации типа «a» возвращает нас к исходному состоянию.

В качестве примера рассмотрим построение топологий для структурной модели показанной на рис.2. Ниже приведены ранговая матрица и схемы построения двух вариантов топологии:

$$R = \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix}, \quad \begin{vmatrix} (12) & 3 \\ 4 & (56) \end{vmatrix} \begin{matrix} (123) \rightarrow V_1^1 \\ (456) \rightarrow V_2^1 \end{matrix} \quad \begin{vmatrix} (12) & 3 \\ 4 & (56) \end{vmatrix} \begin{matrix} (123) \rightarrow V_1^1 \\ (456) \rightarrow V_2^1 \end{matrix}$$

$$\downarrow \quad \downarrow \quad \cdot \quad \downarrow q^B \quad \downarrow q^B$$

$$\begin{matrix} (124) & (356) \\ U_1^2 & U_2^2 \end{matrix} \quad \begin{matrix} (456) & (123) \\ U_1^2 & U_2^2 \end{matrix}$$

Подмножества U_i^1, V_j^2 можно выбрать произвольно. Выберем их равными $U_1^1 = (123), U_2^1 = (456), V_1^2 = (123), V_2^2 = (45)$, в этом случае топологические матрицы первого варианта точно соответствуют представлению (5). Для второго варианта выбрано $q^A = e$, $q^B = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 6 & 1 & 4 & 2 & 3 \end{pmatrix}$ и сохранены прежние значения U_i^1, V_j^2 . Расширенная топология для этого случая имеет следующее матричное представление:

1	1	1								1					1	1
1	1	1									1					1
1	1	1							1							1
			1	1	1						1				1	1
			1	1	1					1					1	1
			1	1	1						1				1	1

От расширенной топологии всегда можно перейти к компактной, выполнив умножение левой или правой смежной топологической матрицы на перестановочную, очевидно, что при этом информация о подстановке q^m теряется и поэтому обратный переход будет неопределен.

В данной работе представлены два уровня структурного описания БНС: структурная модель и топологическая реализация. Показано, что адекватным инструментом проектирования топологий является математический аппарат числовых частичных отображений. На основе

этого аппарата предложен машинно-ориентированный алгоритм построения топологий БНС. Топологическое проектирование позволяет оптимальным образом выбрать программную или аппаратную реализацию БНС из класса эквивалентных.

Литература

- [1] Дорогов А.Ю., Алексеев А.А. Структурные модели быстрых нейронных сетей. В сб. «Интеллектуальные системы» /Труды II-го Международного симпозиума, под ред. К.А.Пупкова, т.2 - М.: Из-во ПАИМС. 1996, с.138-143.
- [2] Дорогов А.Ю., Алексеев А.А. Математические модели быстрых нейронных сетей. В сб. научн. тр. СПбГЭТУ «Системы управления и обработки информации». Вып.490, 1996, с.79-84.
- [3].А.И. Кострикин, Ю.М. Манин. Линейная алгебра и геометрия.- М.: - “Наука” - 1986.-304с.
- [4] Мальцев А.И. Алгебраические системы.- М.: Наука, 1970.
- [5] Кострикин А.И. Введение в алгебру. Основы алгебры. - М.: Физматлит. 1994, -320с.