

УДК 004.383.3

DOI: [10.26102/2310-6018/2020.28.1.023](https://doi.org/10.26102/2310-6018/2020.28.1.023)

Оценка характеристик хранения и функционирования искусственных нейронных сетей на базе нейрокомпьютерного устройства

В.А. Романчук

Федеральное государственное бюджетное образовательное учреждение высшего образования «Рязанский государственный университет имени С.А. Есенина», Рязань, Российская Федерация

Резюме: Актуальность исследования обусловлена недостаточно развитым математическим аппаратом для описания и анализа функционирования класса нейрокомпьютерных устройств при решении задач, представленных в нейросетевом логическом базисе. В связи с этим, данная статья направлена на решение задачи оценки некоторых временных и количественных характеристик хранения и обработки искусственных нейронных сетей на базе нейрокомпьютерных устройств. Основным подходом к решению задачи является подход, согласно которому искусственная нейронная сеть рассмотрена с использованием теоретико-множественного подхода, что позволило получить аналитические соотношения исходя из количества операций эмуляции нейрона и связей между нейронами в соответствии с топологией нейросети. В статье представлены наиболее популярные и используемые топологии искусственных нейронных сетей (однослойный и многослойный перцептрон, сети Хопфилда, сети Хемминга, сети ВАРМ, сети Джордана, сети Элмана, сети АРТ различных модификаций, звезда Гроссберга, сети Кохонена, радиально-базисные нейросети, сети встречного распространения, сверточные сети), для которых приведены временные и количественные аналитические соотношения, позволяющие оценить цикл обучения искусственной нейронной сети, объем необходимой памяти и объем передаваемых данных. Материалы статьи представляют практическую ценность для решения задачи оценки и последующего повышения эффективности функционирования нейрокомпьютерных устройств при решении задач, представленных в нейросетевом логическом базисе, а также для выбора наиболее рациональной модели нейрокомпьютерного устройства.

Ключевые слова: искусственная нейронная сеть, нейропроцессор, хранение, оценка.

Для цитирования: Романчук В.А. Оценка характеристик хранения и функционирования искусственных нейронных сетей на базе нейрокомпьютерного устройства. *Моделирование, оптимизация и информационные технологии*. 2020;8(1). Доступно по: https://moit.vivt.ru/wp-content/uploads/2020/02/Romanchuk_1_20_1.pdf DOI: 10.26102/2310-6018/2020.28.1.023

Evaluation of storage and functioning characteristics of artificial neural networks on the basis of a neurocomputer device

V.A. Romanchuk

Federal State Budgetary Educational Institution of Higher Education "Ryazan State University named for S. Esenin", Ryazan, Russian Federation

Abstract: A solution is proposed for assessing the temporal and quantitative characteristics of the storage and processing of artificial neural networks based on a neurocomputer device. The most popular and used topologies of artificial neural networks are considered (single-layer and multi-layer perceptron, Hopfield networks, Hamming networks, BAM networks, Jordan networks, Elman networks, ART networks of various modifications, Grossberg star, Kohonen networks, radial basis neural networks,

backward propagation networks, convolutional networks) for which analytical relationships are given to evaluate the training cycle of an artificial neural network, the amount of necessary memory and the amount of data transmitted. The difference between the proposed results is that to assess the functioning of artificial neural networks, approaches and characteristics inherent in the class of neuroprocessor devices are offered and only when implementing the calculations presented in the neural network logical basis, which allows to increase the efficiency of the task solution based on neurocomputer devices. An artificial neural network was considered using a set-theoretic approach, which allowed us to obtain analytical relationships based on the number of neuron emulation operations and connections between neurons in accordance with the topology of the neural network.

Keywords: artificial neural network, neuroprocessor, storage, evaluation.

For citation: Romanchuk V.A. Assessment of the characteristics of storage and functioning of artificial neural networks based on a neurocomputer device. Modeling, optimization and information technology.2020;8(1). Available at: https://moit.vivt.ru/wp-content/uploads/2020/02/Romanchuk_1_20_1.pdf DOI: 10.26102/2310-6018/2020.28.1.023 (In Russ).

Введение

Нейрокомпьютинг - это научное направление, занимающееся разработкой и функционированием вычислительных устройств и систем нового, 6-го поколения – нейрокомпьютеров (НК), особенностью которых является большое число параллельно работающих простых вычислительных элементов в составе - искусственных нейронов, выполняющих операции сложения, умножения и функцию активации. Искусственные нейроны связаны между собой так, что составляют искусственную нейронную сеть (ИНС). Они выполняют единообразные вычислительные операции и не требуют внешнего управления. Такая конструкция, содержащая в своем составе множество параллельно работающих вычислительных элементов позволяет добиваться высокого быстродействия [1].

В настоящее время разработка нейрокомпьютеров ведется в большинстве промышленно развитых стран, можно выделить следующие компании: Qualcomm, IBM, HRL Laboratories, Toshiba, HumanBrainProject, KnuEdgeInc., AnalogDevices, TexasInstruments, Darwin, Google, NVidia, Fujitsu, Eyeriss, Intel. В России можно выделить компании: научно-технический центр «Модуль», представивший линейку нейропроцессоров собственной архитектуры NeuroMatrix и компанию «Мотив», разрабатывающую нейрочип «Алтай»,

Нейрокомпьютеры не раз использовались для решения задач машинного обучения, других интеллектуальных задач и зарекомендовали себя как компьютеры, работающие с большой эффективностью. Они прекрасно справляются с задачами распознавания образов на изображениях и видео, управления, поддержки принятия решений, прогнозирования, диагностики и т.д. [2].

При использовании множества нейрокомпьютерных устройств существует проблема: недостаточно развитый математический аппарат для описания и анализа функционирования класса нейрокомпьютерных устройств и, следовательно, при решении задач, представленных в нейросетевом логическом базисе.

Целью исследования является разработка математического аппарата оценки некоторых временных и количественных характеристик хранения и функционирования распространенных искусственных нейронных сетей при реализации их на базе нейрокомпьютерных устройств различной архитектуры.

Материалы и методы

Пусть $Z_{INS}^{(j)}$ - некоторый j -й класс нейросетевых задач, представляющий собой ИНС и который реализован на нейропроцессорном устройстве. Такая задача может быть представлена в виде кортежа параметров и характеристик [3]:

- множество входов ИНС $Net_X = \{net_{x_1}, \dots, net_{x_i}, \dots, net_{x_m}\}$, где каждый вход net_{x_i} характеризуется типом и диапазоном возможных значений;
 - множество выходов ИНС $Net_Y = \{net_{y_1}, \dots, net_{y_i}, \dots, net_{y_m}\}$, где каждый выход net_{y_i} характеризуется типом и диапазоном возможных значений;
 - множество нейронов $Net_N = \{n_1, \dots, n_i, \dots, n_c\}$, каждый из которых необходимо эмулировать для решения некоторого класса задач $Z_{INS}^{(j)}$;
 - множество весовых коэффициентов $Net_W = \{net_{w_1}, \dots, net_{w_i}, \dots, net_{w_{nw}}\}$, характеризующееся типом и диапазоном возможных значений;
 - топология ИНС $Net_T = \{SLP, MLP, RBF, Hop.Ham, BAM, RMLP, Elm, ART, CPN, \dots\}$, включающая число слоев Sl и число нейронов в каждом из слоев;
 - функция активации нейронов F ;
 - способ задания весовых коэффициентов;
- и другие параметры.

В настоящее время используются критерии оценки систем, где система – ИНС на базе нейрокомпьютерного устройства, которые отражают два основных методологических подхода. В первом случае исследуемая система рассматривается как часть более крупной системы, в рамках которой анализируются ее значения, то есть функции входа, выхода. Этот подход подразумевает анализ функционирования исследуемой системы вместе с крупной системой, анализ ее функционального значения, исследование функциональных связей. В этом случае интересующим нас параметром эффективности системы, в первую очередь, является необходимое время для взаимодействия пользователя с системой, в которое входит время обращения к ресурсам, время расчетов, время простоя и др. [5]

Во втором случае рассматриваются исключительно параметры исследуемой системы. При этом характеристики ее связей с более крупной системой выступают в роли ограничений. Рассматривая эффективность в этом случае, нас интересуют такие характеристики, как быстродействие, коэффициент готовности, коэффициент загруженности и т.д.

При исследовании сложных структур - ИНС на базе нейропроцессорных устройств, наиболее рационально использовать характеристики исходя из первого подхода к оценке эффективности системы.

Для реализации задачи хранения обученной ИНС необходимо решить две задачи [6, 7]:

1. Хранение архитектуры ИНС: хранение архитектуры и хранение весовых коэффициентов. Самым важным параметром является объем памяти для хранения весовых коэффициентов для различных топологий ИНС.

2. Обработка ИНС, где важными параметрами являются временная характеристика – цикл обучения ИНС и количественная характеристика – объем передаваемых данных.

Оценка выбранных характеристик для различных топологий ИНС

1. Топология перцептрона

Перцептрон – одна из простейших ИНС, предназначенная для классификации входных векторов с помощью настройки весовых коэффициентов. Устройство однослойного перцептрона (SLP) основано на единственном слое, содержащем nc нейронов [1].

Для данной топологии имеем множества:

- множество входов $Net_X = \{net_{x_1}, \dots, net_{x_i}, \dots, net_{x_n}\}$;
- множество выходов $Net_Y = \{net_{y_1}, \dots, net_{y_i}, \dots, net_{y_n}\}$;
- множество нейронов $Net_N = \{net_{n_1}, \dots, net_{n_i}, \dots, net_{n_c}\}$;
- множество весовых коэффициентов $Net_W = \{net_{w_1}, \dots, net_{w_i}, \dots, net_{w_{nw}}\}$;

Цикл обучения ИНС равен:

$$K_T(SLP) \leq \frac{xn * nc}{H_{CUPC}} + \max_{l \in L} \left(\frac{nc_l * K_{Rl}}{H_{MAC} * H_m} \right), \forall l = \overline{1, L}, \quad \text{где } H_{MAC} - \text{ количество}$$

эмулируемых в НК нейронов за один такт, H_m - количество тактов процессора НК за секунду, K_R - коэффициент загрузки ядра процессора НК.

Необходимый объем памяти можно рассчитать следующим образом:

$$K_O(SLP) = \log_2(\max(net_{wi})) * nw = \log_2(\max(net_{wi})) * xn * nc.$$

Суммарный объем передаваемых данных равен:

$$K_B(SLP) = \sum_{i=1}^{xn} \log_2(\max(net_{xi})) + \sum_{l=1}^{2^{*q}-1} \sum_{j=1}^{nc} \log_2(\max(f_j(\sum_{i=1}^n w_i x_i + w_0))).$$

2. Топология многослойного перцептрона

Многослойный перцептрон (MLP) отличается от классического тем, что имеет более одного слоя нейронов и разделяется на два типа: многослойный перцептрон по Розенблатту и многослойный перцептрон по Румельхарту. Многослойный перцептрон по Румельхарту отличается от многослойного перцептрона по Розенблатту тем, что у него есть обратная связь [1]. Пусть Sl - число слоев перцептрона.

Общее число нейронов равно: $nc(MLP) = \sum_{i=1}^{Sl} nc_i$, где nc_i - количество нейронов в i

-м слое.

Цикл обучения ИНС равен:

$$K_T(MLP) \leq \frac{xn * nc + \sum_{i=1}^{Sl-1} nc_i * nc_{i+1}}{H_{CUPC}} + Sl * \max_{l \in L} \left(\frac{nc_l * K_{Rl}}{H_{MAC} * H_m} \right), \forall l = \overline{1, L}.$$

Необходимый объем памяти можно рассчитать следующим образом:

$$K_O(MLP) = \log_2(\max(net_{wi})) * (xn * nc_1 + \sum_{i=1}^{Sl-1} nc_i * nc_{i+1}).$$

Объем передаваемых данных равен:

$$K_B(MLP) = \sum_{i=1}^{xn} \log_2(\max(net_{xi})) + \sum_{l=1}^{2^{*Sl}-1} \sum_{j=1}^{nc} \log_2(\max(f_j(\sum_{i=1}^n w_i x_i + w_0))).$$

3. Топология радиально-базисных сетей

В классическом случае, радиальная базисная нейронная сеть (RBF) – это двухслойная сеть без наличия обратных связей, которая содержит скрытый слой радиально симметричных нейронов. Скорость обучения сети RBF гораздо выше, чем у перцептрона, но главным недостатком сети RBF является ухудшение точности аппроксимации [1]. Сеть RBF используется обычно для кластеризации объектов исходя из некоторых признаков.

Общее число нейронов равно: $nc(RBF) = nc_1 + nc_2$.

Общее число синапсов равно: $nw(RBF) = xn * nc_1 + nc_1 * nc_2$.

Цикл обучения ИНС равен:

$$K_T(RBF) = \frac{xn * nc + nc_1 * nc_2}{H_{CUPC}} + 2 * \max_{l \in L} \left(\frac{nc_l * K_{Rl}}{H_{MAC} * H_m} \right), \forall l = 1, 2.$$

Необходимый объем памяти можно рассчитать следующим образом:

$K_O(RBF) = \log_2(\max(net_{w1})) * (xn * nc_1) + \log_2(\max(net_{w2})) * (nc_1 * nc_2)$, где net_{w1} - весовые коэффициенты нейронов скрытого слоя, net_{w2} - весовые коэффициенты нейронов выходного слоя.

Объем передаваемых данных равен:

$$K_B(RBF) = \sum_{i=1}^{xn} \log_2(\max(net_{xi})) + \sum_{l=1}^3 \sum_{j=1}^{nc} \log_2(\max(f_j(\sum_{i=1}^n w_i x_i + w_0))).$$

4. Топология Хопфилда

Сеть Хопфилда является рекуррентной ИНС. Особенностью сети Хопфилда являются: наличие обратных связей и блоков динамической задержки, в связи с чем сети часто используются для обработки динамических моделей [1].

Число входов зависит от числа входных параметров и равно числу нейронов и числу выходов: $nc(HopN) = xn(HopN) = yn(HopN)$.

Общее число синапсов равно: $nw(HopN) = xn * nc_1 = xn^2$.

Значения входного, выходного множеств и нейронов могут принимать значения только "1" или "-1":

$$Net_X = \{net_{x1}, \dots, net_{xi}, \dots, net_{xn}\}; net_{xi} \in \{1, -1\}; i = \overline{1, xn},$$

$$Net_Y = \{net_{y1}, \dots, net_{yi}, \dots, net_{yn}\}; net_{yi} \in \{1, -1\}; i = \overline{1, xn},$$

$$Net_N = \{net_{n1}, \dots, net_{ni}, \dots, net_{nc}\}; net_{ni} \in \{1, -1\}; i = \overline{1, xn}.$$

Время обучения ИНС равно: $K_T(HopN) = \frac{xn^2}{H_{CUPC}}$.

Необходимый объем памяти можно рассчитать следующим образом:

$$K_O(HopN) = \log_2(\max(net_{wi})) * nw = \log_2(\max(net_{wi})) * xn^2.$$

Объем передаваемых данных равен:

$$K_B(HopN) = xn + R * 2 * q - 1.$$

5. Топология Хемминга

Сеть Хэмминга состоит из двух слоев, имеющих число нейронов, равное числу образцов. Нейроны первого слоя имеют $xn * xn$ синапсов, соединенных с входами сети. Нейроны второго слоя связаны между собой обратными связями. Единственный синапс с положительной обратной связью для каждого нейрона соединен с его аксоном [1].

Число нейронов для сети Хемминга равно: $nc(Ham) = 3 * xn$.

Общее число синапсов равно: $nw(Ham) = 3 * xn^2$.

Весовые коэффициенты могут принимать значения: $nw(Ham) \in (0, \frac{1}{xn})$.

Количество нейронов в каждом слое одинаково: $nc_1 = nc_2 = nc_3 = xn$.

Время обучения ИНС равно: $K_T(Ham) = \frac{3 * xn^2}{H_{CUPC}}$.

Необходимый объем памяти можно рассчитать следующим образом:

$$K_O(Ham) = \log_2(\max(net_{w1})) * xn^2 + \log_2(\max(net_{w2, w3})) * 2 * xn^2.$$

Объем передаваемых данных равен:

$$K_B(Ham) = \sum_{i=1}^{xn} \log_2(\max(net_{xi})) + R * \sum_{l=1}^5 \sum_{j=1}^{xn} \log_2(\max(f_j(\sum_{i=1}^n w_i x_i + w_0))).$$

6. Топология ВАМ

Двунаправленная ассоциативная память (сеть ВАМ) предназначена для построения реализации возможности ассоциаций каких-либо объектов (образов) с другими; данные в виде вектора поступают на один входной слой нейронов, а соответствующие выходные данные вырабатываются на другом кортеже нейронов.

Общее число нейронов равно: $nc(ВАМ) = 2 * xn$.

Общее число синапсов с учетом двунаправленных связей равно: $nw(ВАМ) = 2 * xn^2$.

Значения входного, выходного множеств и нейронов могут принимать значения только "1" или "-1": $Net_X = \{net_{x1}, \dots, net_{xi}, \dots, net_{xn}\}; net_{xi} \in \{1, -1\}; i = \overline{1, xn}$.

$$\text{Цикл обучения ИНС равен: } K_T(ВАМ) = \frac{2 * xn^2}{H_{СУРС}}$$

Необходимый объем памяти можно рассчитать следующим образом:

$$K_O(ВАМ) = \log_2(\max(net_{w1})) * 2 * xn^2.$$

Объем передаваемых данных равен:

$$K_B(ВАМ) = \sum_{i=1}^{xn} \log_2(\max(net_{xi})) + R * \sum_{l=1}^3 \sum_{j=1}^{xn} \log_2(\max(f_j(\sum_{i=1}^n w_i x_i + w_0))).$$

7. Топология Джордана

Сеть Джордана (RMLP) - вид нейронных сетей, отличающийся тем, что на вход обычного персептрона подается не только входной вектор, но и выходной вектор с задержкой на один или несколько тактов. Входной сигнал передается в группу рецепторов, связанных с входными данными и далее передается в скрытый слой, где, преобразованный скрытым слоем, передается на выходной слой и выходит из сети, а его копия попадает в оперативную память, где задерживается на какое-то время. На следующем этапе, на входные рецепторы поступает второй вектор и, дополнительно, выходной результат с предыдущего шага из памяти [1].

Общее число нейронов равно:

$$nc(RMLP) = \sum_{i=1}^{Sl} nc_i, \text{ где } nc_i - \text{ количество нейронов в } i\text{-м слое.}$$

Общее число синапсов равно: $nw(RMLP) = (xn + yn) * nc_1 + \sum_{i=1}^{Sl-1} nc_i * nc_{i+1}$.

Цикл обучения ИНС равен:

$$K_T(RMLP) \leq \frac{xn * nc_1 + \sum_{i=1}^{Sl-1} nc_i * nc_{i+1}}{H_{СУРС}} + Sl * R * \max_{l \in L} \left(\frac{nc_l * K_{Rl}}{H_{MAC} * H_m} \right), \forall l = \overline{1, L}.$$

Необходимый объем памяти можно рассчитать следующим образом:

$$K_O(RMLP) = \log_2(\max(net_{wi})) * (xn * nc_1 + \sum_{i=1}^{Sl-1} nc_i * nc_{i+1}).$$

Объем передаваемых данных равен:

$$K_B(RMLP) = \sum_{i=1}^{xn} \log_2(\max(net_{xi})) + R * \sum_{l=1}^{2 * Sl-1} \sum_{j=1}^{nc} \log_2(\max(f_j(\sum_{i=1}^n w_i x_i + w_0))) + R * \sum_{i=1}^{yn} \log_2(\max(net_{yi}))$$

8. Топология Элмана

Нейронная сеть Элмана отличается тем, что обратные связи в многослойном

персептроне вводятся не от выхода ИНС, а от выходов скрытого слоя нейронов, что позволяет хранить историю процессов и, чаще всего, более точно решать задачи управления и прогнозирования.

Общее число нейронов равно:

$$nc(ELM) = \sum_{i=1}^{Sl} nc_i, \text{ где } nc_i - \text{ количество нейронов в } i\text{-м слое.}$$

$$\text{Общее число синапсов равно: } nw(ELM) = (xn + yn) * nc_1 + \sum_{i=1}^{Sl-1} nc_i * nc_{i+1}.$$

Цикл обучения ИНС равен:

$$K_T(ELM) \leq \frac{xn * nc_1 + \sum_{i=1}^{Sl-1} nc_i * nc_{i+1}}{H_{CUPC}} + Sl * Ob * \max_{l \in L} \left(\frac{nc_l * K_{Rl}}{H_{MAC} * H_m} \right), \forall l = \overline{1, L}.$$

Необходимый объем памяти можно рассчитать следующим образом:

$$K_O(ELM) = \log_2(\max(net_{wi})) * (xn * nc_1 + \sum_{i=1}^{Sl-1} nc_i * nc_{i+1}).$$

Объем передаваемых данных равен:

$$K_B(ELM) = \sum_{i=1}^{xn} \log_2(\max(net_{xi})) + Ob * Sl * \sum_{l=1}^{2*Sl-1} \sum_{j=1}^{nc} \log_2(\max(f_j(\sum_{i=1}^n w_i x_i + w_0))) + (Sl - 1) * R * \sum_{i=1}^{yn} \log_2(\max(net_{yi}))$$

9. Топология ART

В основе топологии ART лежит адаптивная резонансная теория. ART-сеть содержит единственный слой нейронов, в котором количество входных образов равно количеству признаков, описывающих объект. Количество выходов не является постоянным, во время работы ART-сети количество выходов возрастает при появлении нового незнакомого входного образа, тем самым формируя новый кластер [1].

В настоящее время разработано несколько видов ART нейросетей. Сеть ART-1 функционирует при входных бинарных векторах; ART-2 предназначена для классификации векторов: бинарных и многомерных.

Для сети ART: $|Net_N| = |Net_Y|$; $nc = yn$.

Количество синапсов равно: $nw = 2 * xn * nc$.

Цикл обучения ИНС равен:

$$K_T(ART) = \frac{2 * xn * nc}{H_{CUPC}} + \max_{l \in L} \left(\frac{nc_l * K_{Rl}}{H_{MAC} * H_m} \right), \forall l = \overline{1, L}.$$

Необходимый объем памяти можно рассчитать следующим образом:

$$K_O(ART) = \log_2(\max(net_{wi})) * 2 * nc * xn.$$

Объем передаваемых данных для сети ART-1 равен:

$$K_B(ART1) = xn + \sum_{j=1}^{nc} \log_2(\max(f_j(\sum_{i=1}^n w_i x_i + w_0))).$$

$$\text{для сети ART-2: } K_B(ART2) = \sum_{i=1}^{xn} \log_2(\max(net_{xi})) + \sum_{j=1}^{nc} \log_2(\max(f_j(\sum_{i=1}^n w_i x_i + w_0))).$$

10. Топология сети Кохонена, звезды Гроссберга, сети встречного распространения

Нейронные сети Кохонена чаще всего используется для решения задач прогнозирования и классификации данных, причем обучаясь без учителя. Кроме этого, сеть Кохонена может использоваться с целью сжатия данных [1].

Общее число синапсов равно: $nw(Koh) = xn * nc$.

Топология в виде звезда Гроссберга похожа на топологию сети Кохонена, представляет собой множество нейронов в форме входной звезды, имеющей X входов и один выход $Y = \{y_1\}$, являющийся взвешенной суммой входов [1].

$$\text{Цикл обучения ИНС равен: } K_T(Gr) = \frac{xn * nc}{H_{CUPC}} + \max_{l \in L} \left(\frac{nc_l * K_{Rl}}{H_{MAC} * H_m} \right), \forall l = \overline{1, L}.$$

Необходимый объем памяти:

$$K_o(Gr) = \log_2(\max(net_{wi})) * xn * nc.$$

Объем передаваемых данных равен:

$$K_B(Gr) = \sum_{i=1}^{xn} \log_2(\max(net_{xi})) + \sum_{l=1}^{2*q-1} \sum_{j=1}^{nc} \log_2(\max(f_j(\sum_{i=1}^n w_i x_i + w_0))).$$

Сеть встречного распространения по сути объединяет две рассмотренные топологии и содержит: слой нейронов Кохонена, слой нейронов Гроссберга.

Общее число нейронов равно: $nc(CPN) = nc_k + nc_{xn}$, где nc_k - количество нейронов в слое Кохонена.

$$\text{Общее число синапсов равно: } nw(CPN) = xn * nc_k + nc_k * xn = 2 * nc_k * xn.$$

$$\text{Цикл обучения ИНС равен: } K_T(CPN) = \frac{2 * xn * nc}{H_{CUPC}} + 2 * \max_{l \in L} \left(\frac{nc_l * K_{Rl}}{H_{MAC} * H_m} \right).$$

Необходимый объем памяти можно рассчитать следующим образом:

$$K_o(CPN) = \log_2(\max(net_{wi})) * 2 * nc_k * xn.$$

Объем передаваемых данных равен:

$$K_B(CPN) = \sum_{i=1}^{xn} \log_2(\max(net_{xi})) + 3 * nc_k + \sum_{l=1}^3 \sum_{j=1}^{nc_g} \log_2(\max(f_j(\sum_{i=1}^n w_i x_i + w_0))).$$

11. Топология сверточной нейронной сети

Сверточная нейронная сеть (CNN) – особая топология ИНС, предназначенная для более эффективного распознавания объектов на изображениях. ИНС данной топологии является одной из основных в линейные нейросети глубокого обучения (DeepLearning). Сверточная нейронная сеть является многослойной, без обратных связей и состоит из чередующихся сверточных слоев и слоев подвыборки. Для обучения CNN обычно используют метод обратного распространения ошибки.

Общее число нейронов равно: $nc(CNN) = nc_s + nc_p + nc_{ps}$, где nc_s - количество нейронов в слое свертки, nc_p - количество нейронов в слое пулинга, nc_{ps} - количество нейронов в полносвязном слое.

Если не рассматривать вариант совместного использования параметров, то:

$nc_s = ((Iw - F + 2P) * S + 1) * ((Ih - F + 2P) * S * 1) * Ic$, где F - размер фильтра слоя свертки, S - шаг, с которым перемещаются фильтры, P - количество заполнений нулями на границе входной картинке.

$$nc_p = nc_s / 2.$$

$$nc_{ps} = \sum_{i=1}^{Sl} nc_i, \text{ где } nc_i - \text{ количество нейронов в } i\text{-м слое.}$$

Общее число синапсов равно:

$$nw(CNN) = (F * F * Ic) * nc_s * 1.5 + xn * nc_1 + \sum_{i=1}^{Sl-1} nc_i * nc_{i+1}.$$

Цикл обучения ИНС равен:

$$K_T(CPN) \leq \frac{3 * xn * nc}{H_{CUPC}} + 3 * \max_{l \in L} \left(\frac{nc_s * K_{R1}}{H_{MAC} * H_m}, \frac{nc_p * K_{R2}}{H_{MAC} * H_m}, \frac{nc_{ps} * K_{R3}}{H_{MAC} * H_m} \right).$$

Необходимый объем памяти можно рассчитать следующим образом:

$$K_o(CPN) = \log_2(\max(net_{wi})) * 2 * (F * F * Ic) * nc_s * 1.5 + xn * nc_1 + \sum_{i=1}^{SI-1} nc_i * nc_{i+1}.$$

Объем передаваемых данных равен:

$$K_B(CPN) = \sum_{i=1}^{xn} \log_2(\max(net_{xi})) + \sum_{j=1}^{nc_s} \log_2(\max(f_j(\sum_{i=1}^n w_i x_i + w_0))) + \\ + \sum_{j=1}^{nc_p} \log_2(\max(f_j(\sum_{i=1}^n w_i x_i + w_0))) + \sum_{j=1}^{nc_{ps}} \log_2(\max(f_j(\sum_{i=1}^n w_i x_i + w_0)))$$

Результаты

В ходе исследования были получены аналитические соотношения для оценки количественных и временных показателей при функционировании искусственных нейронных сетей на базе нейрокомпьютерного устройства: время обучения ИНС, необходимый объем памяти и объем передаваемых данных для различных топологий ИНС: однослойный персептрон, многослойный персептрон, сеть Хопфилда, сеть Хемминга, сеть ВАМ, сеть Джордана, сеть Элмана, сеть АРТ, радиально-базисная ИНС, сеть Кохонена, звезда Гроссберга, сеть встречного распространения, сверточная сеть.

Обсуждение

При реализации искусственных нейронных сетей на базе нейрокомпьютерных устройств часто возникает проблема оценки временных и количественных характеристик. Приведенные в работе аналитические соотношения частично решают эту проблему в части оценки времени обучения ИНС, необходимого объема памяти для хранения весовых коэффициентов и объема передаваемых данных для реализации возможности организации облачного вычислителя и распределенных систем. Также необходимо отметить, что приведенные формулы могут быть в дальнейшем уточнены для каждого из типов архитектур нейропроцессорных устройств.

Заключение

В работе предложены аналитические выражения, описывающие эффективность функционирования нейропроцессорных систем. Приведенные в работе аналитические соотношения частично могут быть полезны при выборе конкретного нейропроцессорного устройства для реализации задачи функционирования искусственной нейронной сети, заданной пользователем.

Благодарности

Исследование выполнено при финансовой поддержке Правительства Рязанской области и РФФИ в рамках научного проекта №18-47-623001 p_мол_a.

ЛИТЕРАТУРА

1. Галушкин А.И. *Нейрокомпьютеры*. Кн. 3: учеб. пособие для вузов. ИПРЖР. 2000.
2. ЗАО НТЦ «Модуль». Сайт компании НТЦ «Модуль». 2020. Доступно по адресу: <https://www.module.ru/directions/iskusstvennyj-intellekt> (дата обращения: 01.04.2020).
3. Романчук В.А. Разработка алгоритмического и программного обеспечения контроля функционирования сложных, распределенных и облачных систем на базе нейропроцессоров. *Моделирование, оптимизация и информационные технологии*. 2015;3(10). Доступно по адресу: <http://moit.vivt.ru/wp->

content/uploads/2015/10/Romanchuk_3_15_1.pdf.

4. Романчук В.А., Ручкин В.Н., Фулин В.А. Проектирование нейропроцессорных систем на основе нечеткой кластеризации. *Вестник РГРТУ*. 2014;4(50-1):87-93.
5. Романчук В.А. Математическое обеспечение оценки производительности систем нейрокомпьютерной обработки информации. *Известия Тульского государственного университета. Технические науки*. 2016;2:242-251.
6. Романчук В.А. Метод организации параллелизма данных в нейропроцессорах. *Известия Тульского государственного университета. Технические науки*. 2018;9:148-157.
7. Романчук В.А. Разработка алгоритма оптимизации задачи кластеризации распределенных вычислительных систем на базе нейропроцессоров. *Информатика и прикладная математика: межвузовский сборник научных трудов*. 2016;22:79-83.

REFERENCES

1. Galushkin A.I. *Neirokomp'yutery*. Kn. 3: ucheb. Posobie dlya vuzov. IPRZhR. 2000.
2. ZAO NTTs «Modul'». Saitkompanii NTTs «Modul'». 2020. Dostupno po adresu: <https://www.module.ru/directions/iskusstvennyj-intellekt> (data obrashcheniya: 01.04.2020).
3. Romanchuk V.A. Razrabotka algoritmicheskogo i programmnoho obespecheniya kontrolya funktsionirovaniya slozhnykh, raspredelennykh i oblachnykh system na baze neiroprotsessorov. *Modelirovanie, optimizatsiya i informatsionnye tekhnologii*. 2015;3(10). Dostupno po adresu: http://moit.vivt.ru/wp-content/uploads/2015/10/Romanchuk_3_15_1.pdf.
4. Romanchuk V.A., Ruchkin V.N., Fulin V.A. Proektirovanie neiroprotsessornykh system na osnove nechetkoi klasterizatsii. *Vestnik RGRТУ*. 2014;4(50-1):87-93.
5. Romanchuk V.A. Matematicheskoe obespechenie otsenki proizvoditel'nosti system neirokomp'yuternoii obrabotki informatsii. *Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskie nauki*. 2016;2:242-251.
6. Romanchuk V.A. Metod organizatsii parallelizma dannykh v neiroprotsessorakh. *Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskie nauki*. 2018;9:148-157.
7. Romanchuk V.A. Razrabotka algoritma optimizatsii zadachi klasterizatsii raspredelennykh vychislitel'nykh system na baze neiroprotsessorov. *Informatika i prikladnaya matematika: mezhvuzovskii sbornik nauchnykh trudov*. 2016;22:79-83.

ИНФОРМАЦИЯ ОБ АВТОРЕ / INFORMATION ABOUT THE AUTHOR

Романчук Виталий Александрович, доктор технических наук, профессор, кафедра информатики, вычислительной техники и методики преподавания информатики, ФГБОУ ВО "Рязанский государственный университет имени С.А. Есенина" Рязань, Российская Федерация.
ORCID: 0000-0002-3905-8846
e-mail: v.romanchuk@365.rsu.edu.ru

Vitaliy A. Romanchuk, Dr. Sci. (Phys.–Math.), Professor, Department Of Informatics, Computer Engineering And Methods Of Teaching Informatics, Federal State Budgetary Educational Institution Of Higher Education "Ryazan State University Named By S.Esenin", Ryazan, Russian Federation