

УДК [004.048+004.074]:519.17

Матвийчук А.В.**Скнар И.И.**

Киевский национальный экономический университет имени Вадима Гетьмана

**ВЫБОР ФОРМЫ ПРЕДСТАВЛЕНИЯ ДАННЫХ
БИОЛОГИЧЕСКИ ПОДОБНОЙ НЕЙРОННОЙ СЕТИ**

В статье проведен сравнительный анализ основных форм математического представления ориентированных графов, что может быть использовано для хранения данных биологически подобной нейронной сети, характеризующейся наличием огромного количества структурных элементов и связей между ними. В результате проведенного анализа сделан вывод о неэффективности использования матричного представления для заданной цели – установлено, что хранение параметров нейронной сети в виде списков значительно эффективнее альтернативных вариантов в разрезе необходимого объема памяти. Таким образом, стало возможным оптимизировать способ хранения графов очень большой размерности, например, при построении биологически подобной нейронной сети, во многих технических и экономических задачах, в задачах коммивояжера с большим количеством вершин и т.д.

Ключевые слова: граф, представление графа, биологически подобная нейронная сеть, нейрон, синапс.

Вступление. История человечества знает много примеров, когда в том или ином виде человек создавал технических помощников. С появлением ЭВМ развитие этой отрасли стало невероятно стремительным. Увы, даже столь мощный инструмент показал свою ограниченность в определенных областях: распознавании образов, принятии решений и т.д. Для преодоления этого ограничения был использован созданный в первой половине XX-го в. инструментальный искусственных нейронных сетей. Несмотря на многочисленную критику (самая известная работа по этому вопросу у М. Минского [1]), искусственные нейронные сети с 70-х гг. непрерывно развиваются.

Вместе с вычислительным потенциалом нейронные сети привнесли в мир науки надежду, что их можно будет в определенный момент использовать и для моделирования мозговых структур, что крайне важно и в медицинских целях, и в задачах управления сложными системами (в том числе социально-экономическими), и для дальнейшего понимания и развития нейросетевой концепции. Самой многообещающей в этом вопросе считается концепция спайкинговых нейронных сетей.

Анализ последних исследований и публикаций. Начиная с 2000-х гг., попыт-

ки моделирования нейронных структур мозга получили новое развитие. В 2004 г. Е. Ижикевич описывает более 20 типов поведения нейронов и более десятка различных моделей спайковых нейронных сетей, создав таким образом унифицированную модель [2, с. 1067]. На основе модели Е. Ижикевича стартуют грандиозные проекты: в 2005 г. Манчестерский проект SpiNNaker [3], в 2008 г. – SyNAPSE [4; 5]; в 2009 г. – симулятор спайковых нейронных сетей CARLsim, который учитывает краткосрочную и долгосрочную синаптическую пластичность и даже 4 разных типа ионных каналов (AMPA, NMDA и 2 GABA) [6; 7]. Каждая модель строится на основе определенной конструкции нейронов и определенного метода их соединения и взаимодействия. Одним из наиболее претенциозных проектов является стартовавший в 2005 г. совместный проект компании IBM и Швейцарского федерального технического института Лозанны (EPFL) – Blue Brain Project, который заключается в компьютерном моделировании неокортекса человека.

Постановка проблемы. Отметим, что человеческий неокортекс состоит из порядка 86 млрд нейронов [8], каждый из которых в среднем имеет 10 тыс. синаптических связей. Моделирование структуры подобного масштаба накладывает

жесткие требования и к вычислительной мощности, и к памяти. Оптимизация таких процессов является критичной для подобных исследований.

Выделение не решенных ранее частей общей проблемы. Одним из не решенных вопросов является выбор способа эффективного хранения данных при работе с нейросетью. Основываясь на приведенной выше информации, а также учитывая авторскую концепцию создания биологически подобной искусственной нейронной сети [9, с. 204], которая может быть использована для моделирования в масштабах мозговых макроструктур, можно сформулировать цель и задачи статьи.

Цель статьи – выбор оптимальной формы хранения информации о структуре биологически подобной нейронной сети.

Задачи статьи – анализ существующих методов, выбор оптимального, а также его модификация или создание нового для оптимизации хранения данных биологически подобной нейронной сети.

Изложение основного материала. Для решения поставленных задач прежде всего необходимо определиться с типами информации, которая подлежит хранению. Во-первых, это информация, определяющая межнейронные (синаптические) связи. Это необходимо для четкой идентификации каждого нейрона и его взаимодействия с другими нейронами. В биологическом мире два нейрона могут иметь между собой несколько связей. При моделировании можно упростить до использования одной связи с изменяемой силой. Это вполне допустимо, поскольку появление дополнительных связей обусловлено биологической ограниченностью пропускной способности синапсов в передаче нейромедиаторов в связи с достаточно малой поверхностью пресинаптических и постсинаптических мембран.

Таким образом, во-вторых, нам необходимо хранить информацию о синаптических связях, например о их весовых коэффициентах или пропускной способности и т.д. В-третьих, должна существовать возможность сделать неактивным существующий синапс или же создать новый синапс. Каждый хранимый параметр можно представить самым различ-

ным образом – хранить в виде отдельных переменных, массивов, подмассивов, матриц и т.д. Чтобы выбрать оптимальный способ хранения данных, следует рассмотреть суть представления нейронной сети.

Наиболее наглядным образом нейронную сеть можно представить в виде ориентированного графа, где вершины отображают нейроны, а ребра – синаптические связи. Но стоит помнить, что в отличие от искусственных нейронных сетей (например перцептрона), биологические нейронные сети имеют относительно небольшое количество связей каждого нейрона в сравнении с общим числом нейронов, т.е. их можно представить в виде малосвязного графа. Биологические нейронные сети представляют собой ориентированные графы, где информация по каждому ребру (аксону) передается только в одну сторону. Поэтому будут рассмотрены соответствующие формы представления.

Определим основные понятия. Граф G – пара (V, E) , где $V = \{v_1, v_2, \dots, v_n\}$ – множество вершин, а $E = \{(u, v) | u \in V, v \in V\}$ – множество ребер графа. Если множество ребер состоит из неупорядоченных пар (т.е. $(u, v) = (v, u)$), то граф называется неориентированным, а если пары (u, v) упорядоченные (т.е. в общем случае $(u, v) \neq (v, u)$), то граф называется ориентированным или орграфом. В таком графе ребра принято называть дугами, а граф обозначать \overline{G} [10]. Среди известных математических способов представления графов можно выделить списки ребер и смежности, матрицы инцидентности и смежности.

Матрица инцидентности ориентированного графа $\overline{G}(V, E)$ – матрица $B(\overline{G})$ порядка $n \times m$ ($n=|V|$, $m=|E|$), элементы которой определяются так: $b_{ij} = 1$, если v_i – начало дуги e_j , $b_{ij} = -1$, если v_i – конец дуги e_j ; $b_{ij} = 0$, если v_i и e_j не инцидентны [10].

Размер матрицы $B(\overline{G})$ в памяти можно определить по формуле:

$$M_{\text{inc}} = |V| * |E| * \log_2 3, \quad (1)$$

где выражение $\log_2 3$ (округляется до целого в большую сторону) используется для оценки занимаемой матрицей памяти в двоичном виде, т.к. хранимыми значениями могут быть $\{-1; 0; 1\}$, т.е. три возможных значения.

Для задачи хранения нейросети, состоящей из почти 100 млрд нейронов ($|V| \approx 10^{11}$), имеющих в среднем по 10 тыс. синаптических связей, т.е. $|E| \approx 10^{15}$, размер матрицы $V(\bar{G})$ в памяти составит $\approx 2 \cdot 10^{26}$ bit или же $\approx 20,68$ YB (1 YB = 2^{80} byte).

Нужно учитывать, что при использовании матрицы инцидентности для хранения синаптических связей в памяти, подавляющая часть матрицы будет представлена нулями. Определим количество нулей в матрице инцидентности:

$$M_{incZero} = |V| * |E| - 2 * |E|, \quad (2)$$

где $2 * |E|$ – это суммарное количество исходящих и входящих вершин всех дуг графа \bar{G} , что при $|V| = 10^{11}$ и $|E| = 10^{15}$ в случае с синаптическими связями даст 99,99% от всей матрицы, а это говорит о крайней неэффективности использования такой матрицы.

Более эффективной с точки зрения хранения информации является матрица смежности. Матрица смежности ориентированного графа $\bar{G}(V, E)$, – квадратная матрица $A(\bar{G})$ порядка n ($n = |V|$), элементы которой определяются так: a_{ij} равен числу дуг, ведущих из вершины v_j в вершину v_i (в случае с взвешенным графом), или $a_{ij} = 1$ при наличии дуги из вершины v_j в вершину v_i и $a_{ij} = 0$ при ее отсутствии (для невзвешенного графа) [10].

Матрица смежности по определению имеет меньшую размерность по сравнению с матрицей инцидентности. Также хранимая информация представлена только единицами и нулями даже для направленного графа (2 возможных хранимых значения).

Таким образом, размер матрицы смежности в памяти можно определить по формуле:

$$M_{sm} = |V|^2 * \log_2 2, \quad (3)$$

что для задачи хранения биологически подобной нейросети говорит о необходимости $\approx 1.00114 \cdot 10^{22}$ bit или $\approx 1,06$ ZB (1 ZB = 2^{70} byte) памяти для одной матрицы. То есть использование матрицы смежности вместо матрицы инцидентности экономичнее в 20 тыс. раз.

Для малосвязного графа, как в случае с представлением синаптических связей, большая часть матрицы будет также представлена нулевыми значениями.

Определим размер матрицы, который хранит нули:

$$M_{smZero} = |V|^2 - |E|, \quad (4)$$

где $|E|$ – общее количество хранимых синаптических связей.

Таким образом, большая часть матрицы (99,99%) снова представлена нулями.

Учитывая существенно меньший объем требуемой памяти, можно сделать вывод, что использование матрицы смежности является более эффективным решением, чем матрицы инцидентности.

Список смежности $Ssm(\bar{G})$ – это список, где номеру каждой строки соответствует исходящая вершина, а в строке записаны порядковые номера всех смежных вершин.

Если представить, что максимально возможное хранимое значение равно общему количеству нейронов (вершин), то объем памяти, используемый для хранения списка смежности, будет равен:

$$Ssm = \sum_{i=1}^{|V|} |E_i| * \log_2 |V|, \quad (5)$$

где $|E_i|$ – количество выходных синапсов (связей) i -го нейрона (вершины).

Если $|E_i|$ принять равным в среднем 10^4 , то список смежности займет в памяти $\approx 3,7 \cdot 10^{16}$ bit или $\approx 4,11$ PB (1 PB = 2^{50} byte), что в 270 тыс. раз меньше, чем для матрицы смежности.

Список смежности имеет компактное представление в памяти, но также обладает существенным недостатком – неодинаковой длиной строк. Это существенно усложняет работу с данными, к примеру, в случае, когда необходимо найти вершину орграфа, из которой направлена связь по известной вершине, в которую эта связь входит.

Решить эту проблему возможно, например, создав дубликат списка смежности, но с обратным представлением данных. Так, если первый список по номеру строки будет кодировать вершину с исходящими связями, а в строку записывать перечень вершин, куда эти связи направлены, то второй список будет по номеру строки кодировать вершину с входящими связями, а в строку записывать вершины, из которых эти связи направлены. Таким образом, один список может быть получен из другого простым преобразованием. Но хранение двух списков хотя и удваивает

необходимый объем памяти, при этом обеспечивает более быструю и удобную работу с графом (нейросетью), тем самым повышая скорость работы с данными.

Список ребер графа $Sr(\bar{G})$ – это список, в каждой строке которого записаны две смежные вершины. В отличие от списка смежности, в списке ребер номера нейронов, из которых выходит связь, могут повторяться множество раз – если у вершины больше одного исходящего ребра (а для биологических нейронов в среднем таких соединений больше десяти тысяч), то для хранения каждого последующего ребра необходимо заново записать номер определенной вершины. Если представить, что максимальное хранимое значение равно общему количеству нейронов (вершин), то объем памяти, используемый для хранения списка ребер, будет равен:

$$Sr = 2 * |E| * \log_2 |V|. \quad (6)$$

То есть фактически список ребер графа требует ровно вдвое больше памяти, чем список смежности. Но ранее мы уже упоминали, что для задачи хранения информации о синаптических соединениях и нейронах удобно создавать дополнительный список смежности. А значит, оба решения (2 списка смежности и список ребер) будут занимать равнозначный объем в памяти и оба решения можно считать эффективными. Выбор будет зависеть скорее от удобства использования определенного варианта для решения конкретной задачи.

Иная ситуация складывается, когда для функционирования нейросети необходимо хранить не только факт наличия связи между определенными нейронами, но и вес связи, например, в случае с перцептроном Розенблатта, или даже несколько значений, необходимых для определения переходной функции, например в случае со спайковыми или биологически подобными нейронными сетями [9, с. 204]. В таком случае матрица инцидентности вместо двух единиц с разным знаком должна хранить два одинаковых весовых значения, но также с разными знаками (в случае орграфа), определяющими, входящей или исходящей является связь.

Для хранения же нескольких параметров переходной функции по каждому синапсу нужно хранить дополнительную матрицу инцидентности для каждого дополнительного параметра (вещественного или булевого). Таким образом, объем памяти, необходимый для хранения всех матриц инцидентности, вырастает на величину, рассчитанную по формуле (1) для каждого дополнительного параметра.

Матрица смежности может хранить вместо единицы весовое значение синапса, которое также предполагает представление в вещественной форме. Но для каждого дополнительного весового значения необходимо хранить дополнительную матрицу, в результате чего размер матрицы смежности вырастает на величину, рассчитанную по формуле (3) для каждого дополнительного параметра.

Для списка смежности ситуация аналогичная – необходимо создавать дополнительные списки для каждого следующего хранимого весового коэффициента или параметра переходной функции. Причем если мы используем два списка смежности, как упоминалось выше, то и для каждого хранимого весового коэффициента нужно хранить два списка, что будет занимать в памяти удвоенную величину, рассчитанную по формуле (5). Несмотря на это, такой способ представления данных является достаточно удобным и при этом экономичным.

Список ребер графа может хранить весовые значения в любом количестве в виде дополнительных столбцов. Только в случае со списком ребер не возникает необходимости в создании дополнительных матриц или списков. Благодаря этому размер списка ребер возрастает наиболее медленно: каждый дополнительный хранимый параметр будет занимать

$$Srd = |E| * \log_2 |V|, \quad (7)$$

то есть половину от исходного размера списка ребер и половину от памяти, используемой списком смежности для хранения того же объема данных. Таким образом, список ребер можно считать оптимальной формой хранения информации о структуре биологически подобной нейронной сети.

Другой способ решить проблему хранения нескольких параметров переходной функции для одного синапса – хранить в ячейке рассмотренных выше матриц или списков ссылку на вектор (массив), в котором будут храниться все необходимые параметры определенного синапса. В таком случае можно рекомендовать использовать список ребер или список смежности для хранения ссылок, как наиболее компактные методы из всех рассмотренных.

Выводы. Проведенный анализ основных форм математического представления ориентированных графов позволил показать, что при решении задач хранения графов очень большой размерности, например для хранения данных биологически подобной нейронной сети, наиболее эффективными формами являются список смежности и список ребер графа. Проведенная оценка необходимого объема памяти для хранения отдельных параметров биологически подобной нейронной сети в масштабе целого человеческого мозга показывает, что для хранения информации обо всех существующих синапсах мозга необходимо минимум $\approx 4,11$ РВ при использовании одного списка смеж-

ности. Но более удобное представление займет вдвое больше памяти.

Использование матричного представления для такой цели значительно менее эффективно и требует в десятки тысяч раз большего объема памяти.

В случае необходимости хранения многих параметров нейронной сети в вещественном выражении наиболее эффективным решением является список ребер графа, где для хранения дополнительных параметров используются дополнительные столбцы, что минимизирует занимаемый объем памяти. Также эффективным для решения этой задачи может быть метод, при котором матрица или список хранит ссылки на вектора, которые, в свою очередь, хранят множественные параметры синапса.

Практическая ценность предложенной концепции заключается в возможности оптимизировать способ хранения графов очень большой размерности, например при построении биологически подобной нейронной сети, во многих технических и экономических задачах, в задачах коммивояжера с большим количеством вершин и т.д.

Список литературы:

1. Minsky, M., Papert, S., 1972. *Perceptrons: An Introduction to Computational Geometry* (2nd edition with corrections). The MIT Press, Cambridge MA, 258 p.
2. Izhikevich, E., 2004. Which Model to Use for Cortical Spiking Neurons. *IEEE Transactions on Neural Networks*, vol. 15(5), pp. 1 063–1 070.
3. Furber, S., Temple, S., 2007. Neural systems engineering. *J.R. Soc. Interface*, vol. 4, pp. 193–206.
4. SyNAPSE: IBM Cognitive Computing Project. [online] Available at: <<http://researchweb.watson.ibm.com/cognitive-computing>>.
5. HRL Laboratories, LLC. Center for Neural and Emergent Systems. [online] Available at: <http://www.hrl.com/laboratories/cnes/cnes_main.html>.
6. Carlson, K., Beyeler, M., Dutt, N., Krichmar, J., 2014. GPGPU Accelerated Simulation and Parameter Tuning for Neuromorphic Applications. *IEEE Design Automation Conference (ASP-DAC)*. pp. 570–577.
7. Nageswaran, J., Dutt, N., Krichmar, J., Nicolau, A. and Veidenbaum, A., 2009. A configurable simulation environment for the efficient simulation of large-scale spiking neural networks on graphics processors. *Neural Networks*, vol. 22. pp. 791–800.
8. Azevedo, F.A., Azevedo, F.A., Carvalho, L.R., Grinberg, L.T., Farfel, J.M., Ferretti, R.E., Leite, R.E., Jacob Filho, W., Lent, R. and Herculano-Houzel, S., 2009. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J. Comp. Neurol.*, 513, pp. 532–541.
9. Скнар И.И. Концепция построения биологически правдоподобной искусственной нейронной сети / Скнар И.И. // *Нейро-неіткі технології моделювання в економіці*. – 2014. – № 3. – С. 188–218.
10. *Элементы теории графов : учебное пособие.* / [В.И. Носов, Т.В. Бернштейн, Н.В. Носкова, Т.В.Храмова]. – Новосибирск, 2008. – 107 с.

Матвійчук А.В.

Скнар І.І.

Київський національний економічний університет імені Вадима Гетьмана

ВИБІР ФОРМИ ПРЕДСТАВЛЕННЯ ДАНИХ НЕЙРОННОЇ МЕРЕЖІ, ЩО ПОДІБНА ДО БІОЛОГІЧНОЇ

Резюме

У статті проведено порівняльний аналіз основних форм математичного представлення орієнтованих графів, що може бути використано для збереження даних штучної нейронної мережі, подібної до біологічної, яка характеризується наявністю величезної кількості базових структурних елементів і зв'язків між ними. В результаті проведеного аналізу зроблено висновок про неефективність використання матричного представлення для заданої мети – встановлено, що зберігання параметрів нейронної мережі у вигляді списків значно ефективніше, ніж альтернативні варіанти в розрізі необхідного об'єму пам'яті. У такий спосіб стало можливим оптимізувати збереження графів дуже великої розмірності, наприклад при побудові нейронної мережі, що подібна до біологічної, при вирішенні багатьох технічних і економічних завдань, у завданнях комівояжера з великою кількістю вершин і т.д.

Ключові слова: граф, представлення графа, біологічно подібна нейронна мережа, нейрон, синапс.

Matviychuk A.V.

Sknar I.I.

Kyiv National Economic University named after V. Hetman

CHOOSING OF THE REPRESENTATION FORM FOR THE DATA OF BIOLOGICALLY INSPIRED NEURAL NETWORK

Summary

In this article, a comparative analysis of the main forms of mathematical representation of directed graphs, which can be used for storage of biologically inspired neural networks (that is characterized by the presence of a huge number of structural elements and the relationships between them), was performed. The analysis concluded that matrix form of the graph representation is ineffective for a given purpose. The list form of representation is more efficient than the alternatives in the context of the required amount of memory. Thus, it has become possible to optimize the method for storing large-dimension graphs, which could be used for example, for constructing the biologically inspired neural network, for solving many technical and economic problems, in the traveling salesman problem with a large number of vertexes etc.

Keywords: graph, graph representation, biologically inspired neural network, neuron, synapse.