УДК 004.81

А.И. КУРСИН, НТУ "ХПИ",
А.Г. ЮЩЕНКО, НТУ "ХПИ"

ВОЗМОЖНОСТЬ МЕТАФОРИЗАЦИИ В СЕТЯХ ХОПФИЛДА

Раніш було показано, що нейронні мережі, працюючі за принципом злагодженої активності компонентів та прогнозування вхідних даних, могуть демонструвати явища метафоризації – навчання подібного до утворення понять в людський психіці. На основі порівняльного аналізу таких мереж з нейронними мережами хопфілдовського типу робиться висновок, що є підстави для пошуку зазначених явищ і в мережах Хопфілда. Наводяться результати експериментів.

It was grounded in the previous publications that neural networks maintaining coherent activity of its elements and working via anticipation of its subsequent input may exhibit phenomenon of metaphorization, which is a kind of unsupervised learning resembling concept formation in human thinking. Comparative analysis of anticipatory networks and Hopfield-type models, leads to the conclusion that this phenomenon can be reproduced in simpler Hopfield networks as well. That is supported by simulation results.

Введение. В предыдущих работах [1, 2] нами была предложена модель нейронной сети ансамблевого типа, самоорганизация которой основана на подержании согласованной активности своих элементов и предсказании последующих входных данных – предсказательная нейронная сеть (ПНС). Такая сеть на каждом шаге генерирует предсказания относительно следующих фрагментов данных, которые будут поданы на её вход, сверяет это предсказание с реально поступившими данными и модифицирует свою структуру по результатам сравнения с тем, чтобы генерировать более успешные предсказания в будущем.

Было обосновано, что в такой сети можно ожидать явления самообучения, основанные на кооперации нейронных ансамблей, напоминающие процессы понимания в человеческом мышлении, называемые *метафоризацией* [3]. В самом общем смысле, метафоризация – это создание ментальной модели для нового объекта или явления на основе и по подобию ментальной модели, уже существующей для некоторого похожего объекта или явления.

Однако сложность предложенной нейросетевой архитектуры и большое количество её параметров затрудняют реализацию таких сетей. В связи с этим представляется целесообразным исследовать более простые и распространённые нейросетевые модели, в частности, сети хопфилдовского типа [4] на предмет наличия в них искомых свойств. Особое внимание мы хотели бы обратить на модели из LIF-нейронов, разрабатываемые научной школой под руководством профессора Амита [5 – 7] и используемые этими исследователями для моделирования процессов рабочей памяти, которые наблюдаются в определённых участках коры головного мозга животных при распознавании образов [8, 9]. Эти модели содержат расширения оригинальных сетей Хопфилда, собственно и подвигшие нас к поиску в этом направлении.

Цель данной статьи – провести сравнительный анализ ПНС с нейронными сетями хопфилдовского типа (НСХ) и обосновать попытки воспроизведения метафоризации в них. Для начала кратко охарактеризуем сравниваемые модели.

Предсказательная нейронная сеть. Это сеть ансамблевого типа. Процесс её функционирования состоит в том, чтобы свести текущее возбуждение на входном слое к активности одного из нейронных ансамблей во внутреннем слое сети. Сработавший нейронный ансамбль через ассоциативные связи передаёт мощный сигнал связанным с ним нейронам сети, подготавливая их активность на следующем шаге распознавания. Тем самым формулируется предсказание относительно данных, которые поступят на вход. То есть каждый шаг распознавания заключается во взаимодействии предсказания, сформулированного на предыдущем шаге, и возбуждения от действительно поступающих данных. Специальные механизмы позволяют сети самообучаться, оценивая качество предсказания и модифицируя структуру внутренних связей с целью улучшения предсказаний в будущем.

К таким механизмам относится, прежде всего, 3-х ступенчатая модель активации нейрона [10], в которой промежуточная ступень означает слабую, недостаточную активацию, возникающую вследствие несогласованной активности нейронов, от которых приходят входные связи на данный нейрон. Это позволяет 3-х ступенчатому нейрону выступать датчиком согласованности активности в сети, т.е., например, соответствия предсказания и поступивших данных.

3-х ступенчатая модель нейрона определяет и особое правило модификации межнейронных связей, по которому усиливаются связи, приводящие нейрон в состояние высокой активности, и ослабляются – те, которые доводят его только до промежуточного уровня. Тем самым в памяти сети закрепляются случаи согласованных внутренних состояний. Дополнительно, для повышения качества работы сети, в механизм обучения вводят постепенность закрепления изменений – в зависимости от продолжительности последовательности успешных предсказаний.

Регулирующая, тормозная функция осуществляется управляющим центром, который регулирует пороги перехода нейронов между состояниями активности, повышая их с ростом числа активных нейронов в сети. Действие центра дополняется ростом усталости отдельных нейронов, когда постепенно исключаются наиболее слабые претенденты из борьбы за высокие уровни активации.

В ПНС нейронный ансамбль, выделяемый за счёт более сильной связности нейронов, функционирует как группа взаимопомощи [10], члены которой помогают друг другу достигать высоких уровней активации. Взаимопомощь возникает и между ансамблями, а именно при обучении новым данным, когда вновь обучаемый ансамбль кооперируется с похожим уже обученным. Такая взаимопомощь поощряется сочетанием неполной связности нейронов сети, 3-х ступенчатой моделью нейрона, пересекающимися ансамблями и необходимостью постоянно генерировать предсказания. При этом новый ансамбль, как правило, более подходящий для новых входных данных, помогает старому, недостаточно подходящему, сработать (довести свои нейроны до высокой степени активации), а старый предоставляет новому свои ассоциативные связи для генерации предсказаний, пока собственные связи новичка не окрепнут в достаточной степени.

Если мы будем рассматривать нейронный ансамбль вместе с его ассоциативными связями как аналог ментальной модели, то получим процесс, при котором новая ментальная модель создаётся на базе старой, что фактически и понимается под *метафоризацией* [3].

Описанная модель нейронной сети позволяет предположить наличие в ней интересных явлений самообучения. Однако она обладает и рядом недостатков. Главный из них – большая сложность и многопараметричность модели (подробное описание см. в [2]). Это, в сочетании с недостаточной экспериментальной проработкой её ближайших аналогов [10 – 12], затрудняет реализацию действующих моделей такой сети. В связи с этим в настоящий момент реализована только самоорганизация сети в холостом режиме – без входных данных – в цепочку последовательных спонтанных срабатываний нейронных ансамблей [2]. Также, остаётся недостаточно разработанным механизм предсказаний более чем на один шаг. В связи с этим, модели, рассматриваемые далее, обладают определёнными преимуществами. Им также присущ аттракторный характер функционирования. Они воспринимают последовательности входных данных и способны к их предсказанию, причём, более чем на один шаг.

Сети Хопфилда. Сеть этого типа [4], как правило, функционирует следующим образом. Периодически на её вход подаются входные данные, вызывающие определённый рисунок возбуждения внутри сети. После отключения входных данных сеть переходит в одно из внутренних состояний, называемых аттракторами. Выбор аттрактора зависит от входных данных и от внутренней структуры сети. Так, если сеть структурирована на нейронные ансамбли, то аттрактором, скорее всего, станет ансамбль, наиболее близкий к входным данным. В состоянии аттрактора сеть может находиться продолжительное время. В этом иногда видят аналогию со способностью человеческого мышления сохранять временную память о внешних объектах или явлениях, когда они уже не воздействуют на органы чувств.

Сети хопфилдовского типа строятся на нейронах разного уровня сложности: от простых бинарных – до LIF-нейронов, достаточно близко моделирующих пирамидальные нейроны коры [6]. Связи в таких сетях обучаются, как правило, хеббовским способом. Например, в [13] предложено интересное правило обучения связей хеббовского типа, которое сочетает долговременную потенциацию и долговременное торможение, а также порождает кратковременные изменения проводимости, служащие основой кратковременной памяти в аттракторах. Простой механизм торможения, увеличивающий тормозной сигнал пропорционально суммарной активности нейронов и подающий его одинаково на все нейроны сети, достаточен для поддержания количества активных нейронов в определённых пределах [5].

Показано, что НСХ обладают спонтанным состоянием активности, возникающим в ответ на пустой (зашумленный) вход – глобальным аттрактором, а также – специфическими аттракторами, которые формируются для фрагментов данных, подаваемых на вход сети в процессе обучения [5]. Как правило, входные данные подаются на сеть путём возбуждения конкретных нейронов на внутреннем и единственном слое сети, а обучению при этом подвергаются ассоциативные связи таких нейронов. Показано, что при определённых процедурах обучения, сеть образовывает внутри себя нейронные ансамбли-аттракторы для фрагментов входных данных, взаимосвязанные между собой ассоциативно в соответствии с протоколами обучения [5, 6].

Сравнительный анализ нейросетевых моделей. Несмотря на отмеченные существенные различия, в описанных моделях нейронных сетей можно усмотреть определённые сходства. Функционирование обоих типов нейронных сетей имеет аттракторную природу. Как в ПНС и НСХ, аттрактором может быть наиболее активный нейронный ансамбль. Кратковременность срабатывания ансамбля в ПНС является в определённой степени минусом по сравнению с его устойчивой продолжительной активностью в НСХ – если принять во внимание способность человеческой психики сохранять рабочую память о внешнем воздействии уже после прекращения последнего. Наличие кратковременной памяти, которая в ПНС распределена: а) по уровням активности нейронов; б) по нескольким компонентам модифицируемых связей; в) по показателям уровня усталости нейронов, а в НСХ поддерживается: а) динамической компонентой проводимости связей и б) продолжительной коллективной активностью нейронов аттрактора. Также описанный выше механизм торможения в НСХ, несомненно, проще и может быть достаточно эффективным по сравнению с торможением при помощи глобальных порогов и нарастающей усталости нейронов, предложенном в ПНС. Различия в подаче входных данных: через тренируемые связи от входного слоя (в ПНС) и непосредственной активацией нейронов – в НСХ могут быть несущественными на начальном этапе моделирования и устранены в дальнейшем.

Можно сделать вывод, что ПНС обладают, несомненно, более мощным потенциалом, например, функции показателя усталости нейрона могут быть использованы для разделения нейронов на фракции по частоте участия в работе сети. Однако большие возможности ПНС делают их и гораздо более сложными для моделирования. Вместе с тем, НСХ по многим параметрам предлагают разумно более простую альтернативу. Это побудило нас рассмотреть сети Хопфилда в качестве альтернативной модели для воплощения метафоризации.

Поиск метафоризации в НСХ. Метафоризация предполагает наличие в сети пересекающихся нейронных ансамблей: трудно предположить, чтобы сходство между входными паттернами не отражалось в общности нейронов для ансамблей, представляющих эти паттерны внутри сети. Пересечение ансамблей происходит не только по нейронам, но и по связям, их объединяющим. Следовательно, можно предположить, что у вновь тренируемого ансамбля, который имеет пересечение с уже обученным, часть связей, находящаяся в этом пересечении, уже обучена должным образом. Этот «аванс» обучаемый ансамбль может использовать для своего более быстрого обучения. В эксперименте, проведенном нами с довольно простой сетью хопфилдовского типа, мы получили подтверждение данной гипотезы. Ансамбли, пересекающиеся с уже обученными, действительно обучаются быстрее (см. рис. 1) и, что самое главное, при соблюдении определённых ограничений на величину пересечения, обученные таким образом ансамбли сохраняют способность срабатывать впоследствии по отдельности.

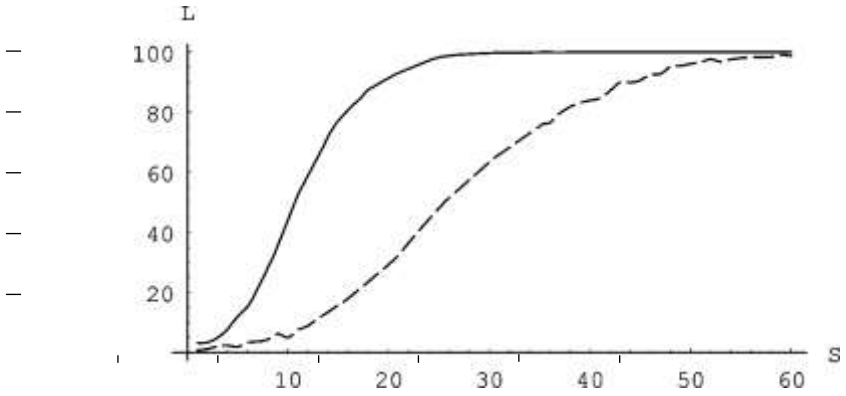

Рис. 1. Зависимость меры обучения L (%) – средней для ансамблей из обучаемого множества – от количества повторов (S) обучающей последовательности. Сплошная линия соответствует ансамблям, имеющим пересечения с ранее обученными, пунктирная – ансамблям, не имеющим таких пересечений.

Подробно эти эксперименты будут описаны позже. Здесь же мы хотим заметить, что полученное ускорение обучения пересекающихся ансамблей вполне можно рассматривать как упрощённое проявление кооперации между ансамблями в процессе метафоризации. Действительно, можно предположить только два способа осуществления такой кооперации: через общие нейроны или посредством ассоциативных связей. Из них – первый является наиболее действенным и вероятным, т.к. ассоциативные связи у нового, ранее не задействованного ансамбля будут, скорее всего, слишком слабы, чтобы обеспечить достаточную кооперативную помощь, хотя и этот аспект необходимо исследовать в полном объёме.

Выводы. Сравнительный анализ структур и функционирования предсказательной нейронной сети и сетей Хопфилда вместе с описанными экспериментами по обучению пересекающихся ансамблей подтвердили правомерность поиска явлений метафоризации в сетях хопфилдовского типа. Ускорение обучения ансамблей за счёт пересечения с уже обученными является несомненным проявлением кооперации между ментальными моделями, которая предполагается при метафоризации. Дальнейший поиск планируется направить на развитие опробованной модели: введение ассоциативных связей между ансамблями, моделирование предсказаний последовательностей данных и зависимости обучения от качества предсказаний, присутствующих в ПНС. Эти исследования, по нашему мнению, могут сделать возможным полное моделирование метафоризации и других сложных когнитивных процессов в сетях хопфилдовского типа.

Благодарности. Данное исследование выполнено при поддержке INTAS, грант YSF 03-55-1661.

Список литературы: 1. *Kursin A.* Neural Network: Input Anticipation May Lead To Advanced Adaptation Properties // Artificial Neural Networks and Neural Information Processing. – Berlin-Heidelberg: Springer-Verlag, 2003. – С. 779–785. **2.** *Kursin A.* Self-Organization of Anticipatory Neural Network // Scientific Proceedings of Riga Technical University, series Computer Science, Information Technology and Management Science. – Riga: RTU. – 2004. – Vol. 20. – С. 51–59. **3.** *Сергеев В.* Проблема понимания: некоторые мысленные эксперименты // Теория и модели знаний. Труды по искусственному интеллекту. Уч. Зап. Тартуского гос. ун-та. – Тарту: Изд-во Тарт. ун-та, 1985. – Вып. 714. – С. 133–147. **4.** *Hopfield J.* Neural Nets and Physical Systems with Emergent Collective Computational Abilities // Proc. of the National Academy of Sciences USA. – 1982. – Vol. 79. – P. 2554–2558. **5.** *Amit D., Brunel N., Tsodyks M.* Correlations of Cortical

Hebbian Reverberations: Experiment Versus Theory // Journal of Neuroscience. – 1994. – № 14. – P. 6435–6445. **6.** *Brunel N.* Hebbian Learning of Context in Recurrent Neural Networks // Neural Computation. – 1996. – № 8. – P. 1677–1710. **7.** *Brunel N.* Dynamics and Plasticity of Stimulus-Selective Persistent Activity in Cortical Network Models // Cerebral Cortex. – 2003. – № 13. – P. 1151–1161. **8.** *Miyashita Y.* Neuronal Correlate of Visual Associative Long-Term Memory in the Primate Temporal Cortex // Nature. – 1988. – Vol. 335. – P. 817–820. **9.** *Sakai K., Miyashita Y.* Neural Organization for the Long-Term Memory of Paired Associates // Nature. – 1991. – Vol. 354. – P. 152–155. **10.** *Емельянов-Ярославский Л.* Интеллектуальная квазибиологическая система. Индуктивный автомат. – М.: Наука, 1990. – 112 с. **11.** *Wickelgren W.* Webs, Cell Assemblies, and Chunking in Neural Nets // Canadian Journal of Experimental Psychology. – 1999. – Vol. 53. – № 1. – P. 118–131. **12.** *Амосов Н. и др.,* Автоматы и разумное поведение. – К.: Наукова думка, 1973. – 370 с. **13.** *Amit D., Brunel N.* Learning Internal Representations in an Attractor Neural Network // Network. – 1996. – № 6. – P. 359–388.

Поступила в редакцию 14.04.2006