

Измерение параметров порядка на основе нейросетевых технологий

Ю. В. ВОХМИНА, В. М. ЕСЬКОВ, Т. В. ГАВРИЛЕНКО, О. Е. ФИЛАТОВА

Сургутский государственный университет Ханты-Мансийского автономного округа — Югры, Сургут, Россия, e-mail: filatovmik@yandex.ru, kafedra_bin@mail.ru

Продemonстрирована сходимость процедуры многократного обучения нейронной сети, когда при многократном повторении решения задачи бинарной классификации на основе неизменных пакетов обучающих выборок нейромодулятор может демонстрировать первоначально хаотический набор значений весов признаков, который с увеличением числа итераций сводится к некоторому ранжированному ряду.

Ключевые слова: биосистема, параметры порядка, нейросеть, теория хаоса-самоорганизации.

When the problem of binary classification based on an unchanged set of learning samples is solved many times, the neural emulator can exhibit chaotic originally set values of the features weights, which with increased number of iterations converges to a certain number of the ranked features weights.

Key words: biosystem, order parameters, neural network, theory of chaos-self-organization.

Возникновение синергетики связано с возможностью и необходимостью решения проблемы идентификации параметров порядка и создания теории системного синтеза в целом. Несмотря на огромные усилия ученых за последние 40 лет существенных достижений в области формализации процедуры идентификации параметров порядка и всего системного синтеза не было. Идентификация параметров порядка остается уникальной процедурой для отдельных избранных задач. Универсальная формализация системного синтеза отсутствует в настоящее время для сложных систем, к которым, в первую очередь, относятся биосистемы [1, 2].

Вместе с тем, существует определенный класс задач и систем, где такая формализация возможна при применении нейросетевых технологий. В последние годы нейромодуляторы широко используют при идентификации наиболее важных диагностических признаков в медицине и биологии, однако результат, получаемый при их применении, не всегда однозначен, а значит, использовать нейромодуляторы в практических целях не имеет смысла. Существует ли выход из этой ситуации? Ответ представлен в настоящей статье на примере бинарной классификации, т. е. при попытках выделения параметров порядка и ранжирования весов диагностических признаков — компонентов x_i вектора состояния системы $\mathbf{x} = \mathbf{x}(t)$.

Непараметрическое распределение весовых признаков нейромодулятора. Задача бинарной классификации с помощью нейро-ЭВМ очень распространена в медицине, так как позволяет обобщенно, без отдельного анализа динамики изменения каждого компонента x_i вектора состояния организма человека (ВСОЧ) $\mathbf{x} = \mathbf{x}(t) = (x_1, x_2, \dots, x_m)^T$ установить, чем отличается одна группа больных (например, без лечения) от другой группы (с тем или иным видом лечения). На основе бинарной классификации можно сравнивать группы пациентов, получающих разные виды помощи (лекарства, физиотерапию и т. д.) [2, 3], устанавливать эффективность оздоровительных мероприятий и влияние экологических факторов среды, оценивать тренерскую работу в спорте.

После выявления нейромодулятором различий надо оценить значимость (вес) того или иного i -го диагностического признака ($i = 1, 2, \dots, m$, где m — размерность фазового пространства), т. е. компонент ВСОЧ x_i , что особенно важно, когда эти признаки характеризуют состояние разных функциональных систем организма человека. Указания врачу (на что ему следует обратить внимание) необходимы для формального решения задачи бинарной классификации, а определение наиболее значимых диагностических признаков —



Диаграмма распределения весовых коэффициентов каждого из параметров x_i для каждого j -го обучения (метод градиентного спуска) искусственной нейронной сети ($j = 1, \dots, 50$)

важный раздел клинической медицины — для выявления, например, региональных (эндемических) особенностей протекания заболеваний или при переходе к индивидуализированной медицине, когда параметры порядка у каждого больного могут быть разными при одинаковом заболевании [4, 5].

Изучение особенностей протекания заболевания у конкретного пациента или группы больных можно подтвердить при решении задачи бинарной классификации путем ранжирования весов признаков, т. е. каждого x_i . Однако при ха-

отическом задании начальных весов w_{i0} признаков x_i , w_{i0}^j в момент $t = 0$, т. е. до начала настройки нейросети ($j = 1, 2, \dots, p$ — итерации настройки нейронной сети, образующие группу итераций $l = 1, 2, \dots, p$, состоящую из p итераций каждая) и при следующих повторах решения задачи (переходу от k к $k+1$ и т. д.) бинарной классификации, веса признаков w_{ik}^j (k —

вес в конце каждой настройки сети) получают различные. Характерный пример представлен на рисунке, где изобра-

жена диаграмма идентификации конечных весов w_{ik}^j признаков для диагностики различий в состоянии кардиореспираторной функциональной системы (группы из 19 испытуемых, у которых измерялось $m = 13$ параметров ВСОЧ, т. е. x_i , где $i = 1, 2, \dots, 13$ при $p = 50$ повторах решения задачи бинарной классификации, $j = 1, \dots, p$, где $p = 50; 100; 1000$. Здесь x_1 — NN — значение кардиологических интервалов в выборке; x_2 — SpO₂ — концентрация оксигемоглобина в крови; x_3 — SIM — индекс активности симпатического отдела вегетативной нервной системы; x_4 — PAR — индекс активности парасимпатического отдела вегетативной нервной системы; x_5 — SDNN — стандартное отклонение измеряемых кардиологических интервалов; x_6 — HRV — отношение общего числа кардиологических интервалов к количеству интервалов с наиболее часто встречающейся длительностью (амплитуде моды); x_7 — IB — индекс напряжения (по Р. М. Баевскому); x_8 — VLF, x_9 — LF, x_{10} — HF — спектральные мощности очень низких частот, низких и высоких частот, соответственно; x_{11} — LF/HF — отношение низкочастотной составляющей к высокочастотной.

На рисунке показано 50 повторений (итераций настройки нейро-ЭВМ) для каждого из 13 диагностических признаков, не совпадающих в каждой итерации. После 50 итераций получали выборки весов признаков по каждой i -й координате фазового пространства состояний, которые на каждом l -м шаге статистически обрабатывали (находили средние $\langle w_{ik} \rangle$ по всем j для l -й группы итераций и их дисперсии). Выборки не совпадали при малых p , они получались из разных генеральных совокупностей.

Идентификация параметров порядка на основе сходимости весов признаков. Установлено, что для группы из 19 человек (измерения параметров ВСОЧ проводили до работы испытуемых в ночную смену и после) при 50 повторах нейроэмулятор каждый раз разделяет эти группы, но веса w_{ik}^j признаков x_i при этом получают разные и не удовлетворяют закону нормального распределения (получали разные функции распределения). При нарастании числа повторов p обучения нейросети при одинаковых выборках наблюдается сходимость усредненных значений весов признаков по каждому признаку x_i к некоторому среднему значению $\langle w_i^j \rangle$.

Хаотические $\langle w_i^j \rangle$ колебания четко проявляются в двух последних значащих цифрах. Иными словами, если проводятся измерения в пределах двух значащих цифр ($p < 100$), то относительная амплитуда колебаний будет высока. При переходе к $100 < p < 1000$ удерживается первая значащая цифра, а при переходе к $1000 < p < 10000$ — две значащие цифры и т. д. Набор из l выборок по p весов w_i в каждой с увеличением p показывает сходимость получаемых для каждого диагностического признака i , при этом для каждого признака i может быть разной.

В табл. 1, 2 приведены результаты расчета двух типов выборок по $p = 100; 1000$ итераций в каждом l -й группе в режиме бинарной классификации. Параметры w_i — это показатели сердечно-сосудистой системы людей перед их работой в ночную смену и после. Из таблиц следует, что ампли-

туда колебаний средних значений весов признаков $\langle w_i^j \rangle$ при

$p = 100$ значительно больше, чем при $p = 1000$. Представлены также интервалы разброса средних значений весов признаков $\langle w_i \rangle$ по каждой из четырех выборок по $p = 100$ вместе с $p = 400$ (общая выборка). Если по первому признаку интервал изменений составляет $\Delta w_i = 0,121$ для $p = 100$, то для $p = 1000$ интервал разброса уменьшается почти в десять раз и $\Delta w_i = 0,0117$. Расчет проводили до $p = 10^6$, на миллионе итераций наблюдали сходимость разброса к четырем первым значащим цифрам.

Отметим еще одну закономерность: законы распределения при $p < 100$ для каждого набора повторов настроек нейросетей ($p_1 < 100, p_2 < 100, \dots$) не удовлетворяют нормальному закону распределения и каждый раз возникает генеральная совокупность с выборкой и различными функциями распределения. Иными словами, хаос проявляется в малых выборках, но при больших $p > 10^3$ нарастает упорядоченность в работе нейроэмулятора, хотя распределения остаются непараметрическими и могут различаться. Можно предположить, что аналогично работают и нейросети мозга, когда при малом числе ревербераций в гиппокампе человек не получает правильного решения задачи, так как не может долго удерживать в сознании решение задачи.

В процессе эволюции в биосфере тоже могут происходить подобные эффекты: при малых повторах мутаций ($p < 100$) не получается новый эффект (например, возникновение нового свойства организма или нового вида), но с течением времени при нарастании p биологический процесс начинает устойчиво сходиться к первым значащим цифрам. При $p \geq 100$ неизменной будет первая цифра из трех, при $p \geq 1000$ — это уже будут две значащие цифры из четырех и т. д. Подобные проблемы описаны в [4] группой ученых из Стенфорда, которые изучали организацию движений. Можно сделать вывод, что при многократных повторах (тренингах) точность выполнения движений увеличивается [5, 6].

Таким образом, при увеличении числа итераций p получают определенные значения весов признаков, позволяющие ранжировать значения весов w и определять параметры порядка. Точность идентификации значимости конкретного диагностического признака повышается и при переходе к $p \geq 10^6$ параметры порядка можно устойчиво ранжировать.

Таблица 1

Усредненные значения отдельных координат весов признаков w_i вектора состояния системы при идентификации параметров порядка нейроэмулятором для $100 \leq p \leq 400$ итераций

Расчеты итераций по выборкам ($N > 100$)	LF: $\langle w_1 \rangle$	HF: $\langle w_2 \rangle$	SIM: $\langle w_3 \rangle$	NN: $\langle w_{13} \rangle$	SPO ₂ : $\langle w_{14} \rangle$	CSS: $\langle w_{15} \rangle$
$p = 400, j = (1, \dots, 400)$	0,77	0,664	0,629	0,269	0,247	0,231
$p = 100, j = (1, \dots, 100)$	0,831	0,664	0,614	0,260	0,275	0,247
$p = 100, j = (100, \dots, 200)$	0,773	0,683	0,683	0,287	0,247	0,242
$p = 100, j = (200, \dots, 300)$	0,77	0,674	0,646	0,261	0,23	0,223
$p = 100, j = (300, \dots, 400)$	0,71	0,673	0,589	0,264	0,231	0,217
Интервалы изменений $\langle \Delta w_i \rangle$	0,121	0,019	0,094	0,027	0,045	0,025

Таблица 2

Усредненные значения отдельных координат весов признаков w_i вектора состояния системы при идентификации параметров порядка нейроэмулятором для $1000 \leq p \leq 5000$ итераций

Расчеты итераций по выборкам ($N > 1000$)	LF: $\langle w_1 \rangle$	SIM: $\langle w_2 \rangle$	HF: $\langle w_3 \rangle$	CSS: $\langle w_{13} \rangle$	NN: $\langle w_{14} \rangle$	SPO ₂ : $\langle w_{15} \rangle$
$p = 5000, j = (1, \dots, 5000)$	0,8012	0,5655	0,5603	0,4413	0,3709	0,337
$p = 1000, j = (1, \dots, 1000)$	0,7985	0,565	0,557	0,4443	0,3647	0,3337
$p = 1000, j = (1000, \dots, 2000)$	0,8058	0,5663	0,5641	0,4371	0,3642	0,3362
$p = 1000, j = (2000, \dots, 3000)$	0,7957	0,5678	0,558	0,4426	0,3678	0,3339
$p = 1000, j = (3000, \dots, 4000)$	0,8032	0,5571	0,558	0,4353	0,3774	0,3336
$p = 1000, j = (4000, \dots, 5000)$	0,8074	0,5741	0,5666	0,4526	0,3786	0,3536
Интервалы изменений $\langle \Delta w_i \rangle$	0,0117	0,017	0,0096	0,0173	0,0144	0,0166

Выводы. Представлены возможные механизмы принятия решений в виде выбора параметров порядка нейронными сетями мозга человека и животных на основе многократного повторения (итераций) настроек нейросети при хаотическом начальном распределении весовых коэффициентов диагностических признаков. На основе моделей нейросетевых систем, работающих в режимах реверберации, продемонстрирована возможность выбора параметров порядка в задачах бинарной классификации.

При малом числе повторов ($p < 100$) выполнения решающих функций возможность ошибки велика, т. е. нейросеть выдает или неправильное решение, или искаженное, но приближенное к правильному.

Повторение решения одной задачи (например бинарной классификации) имеет некоторую аналогию с организацией постурального тремора и регуляцией кардиологических интервалов [2, 7, 8], когда большие порядки (первые значащие цифры после запятой в значениях весов признаков) остаются неизменными, а при последующих порядках происходит постоянно изменяющиеся непараметрические распределения, которые в общем случае могут иметь разные генеральные совокупности.

Полученные результаты имеют большое теоретическое и практическое значение, в том числе для метода многомерных фазовых пространств, рассматривающих квазиаттракторы — объемы фазовых пространств, внутри которых вектор состояния системы $x_i = x(t) = (x_1, x_2, \dots, x_m)^T$ совершает непрерывное хаотическое движение [1, 2]. В примере с нейроэмулятором в качестве регулируемых переменных выступают веса динамических признаков, влияющие на механиз-

мы задания управляющих воздействий в регуляции тремора и других нестационарных процессов в реальных биологических системах.

Литература

1. Еськов В. М., Еськов В. В., Гавриленко Т. В., Зимин М. И. Неопределенность в квантовой механике и биофизике сложных систем // Вестник Московского университета. Физика. Астрономия. 2014. № 5. С. 41—46.
2. Eskov V. M., Eskov V. V., Filatova O. E., Filatov M. A. Two types of systems and three types of paradigms in systems philosophy and system science // J. Biomedical Science and Engineering. 2012. V. 5. N. 10. P. 602—607.
3. Churchland M. M., Cunningham J. P., Kaufman M. T. Neural population dynamics during reaching // Nature. 2012. V. 487. P. 51—56.
4. Еськов В. М., Кулаев С. В., Попов Ю. М., Филатова О. Е. Компьютерные технологии в измерении устойчивости стационарных состояний в биологических динамических системах // Измерительная техника. 2006. № 1. С. 59—65; Eskov V. M., Kulaev S. V., Popov Yu. M., Filatova O. E. Computer technologies in stability measurements on stationary states in dynamic biological systems // Measurement Techniques. 2006. V. 49. N. 1. P. 59—65.
5. Еськов В. М., Еськов В. В., Филатова О. Е. Особенности измерений и моделирования биосистем в фазовых пространствах состояний // Измерительная техника. 2010. № 12. С. 53—57; Eskov V. M., Eskov V. V., Filatova O. E. Characteristic features of measurements and modeling for biosystems in phase spaces of states // Measurement Techniques. 2010. V. 53. N. 12. P. 1404—1410.

6. Еськов В. М., Гавриленко Т. В., Козлова В. В., Филатов М. А. Измерение параметров динамики микрохаоса в поведении реальных биосистем // Метрология. 2012. № 7. С. 39—48; Eskov V. M., Gavrilenko T. V., Kozlova V. V., Filatov M. A. Measurement of the dynamic parameters of microchaos in the behavior of living biosystems // Measurement Techniques. 2012. V. 55. N. 9. P. 1096—1100.

7. Haken H. Principles of brain functioning: a synergetic approach to brain activity, behavior and cognition. Berlin: Springer, 1996.

8. Prigogine I. The Die Is Not Cast // Futures. Bulletin of the Word Futures Studies Federation. 2000. V. 25. N. 4. P. 17—19.

Дата принятия 08.10.2014 г.

004.932.2

Современные принципы измерений в интеллектуальных системах гистологической диагностики онкологических заболеваний

В. Г. НИКИТАЕВ

Национальный исследовательский ядерный университет «МИФИ»,
Москва, Россия, e-mail: kaf46@mail.ru

Рассмотрены проблемы измерений в интеллектуальных (на базе распознавания образов) системах гистологической диагностики онкологических заболеваний. Обсуждены принципы проведения исследований на этапе микроанализа опухолей, основанные на оценке характеристик тканевых и клеточных структур с применением различных шкал измерений.

Ключевые слова: информационно-измерительные системы, гистологическая диагностика, распознавание образов, онкологические заболевания, морфологические характеристики, шкалы измерений.

The problems of measurement in the intellectual systems (based on pattern recognition) of histological diagnosis of cancer at the present stage are discussed. The approach to the measurement on the stage of microanalysis of tumors is based on the assessment of characteristics of tissue and cellular structures with the use of measurement scales has been proposed.

Key words: information-measuring system, histological diagnosis, pattern recognition, cancer, morphological characteristics, measurement scales.

При создании и практическом применении современных высокотехнологичных интеллектуальных диагностических систем в медицине и технике с использованием методов распознавания изображений разработчики сталкиваются с рядом интуитивно воспринимаемых и неоднозначных понятий. При решении измерительных задач гистологической диагностики (ГД) онкологических заболеваний в число этих понятий входят: граница, длина, цвет, форма, структура изображений опухолей. Изображения гистологических макро- и микропрепаратов опухолей характеризуются стохастической пространственно-яркостной организацией элементов, размытостью границ, сложностью цветовой палитры (свыше 16 млн цветовых оттенков на пиксел), слабоформализуемым морфологическим разнообразием тканевых структур и клеточных форм. На результаты измерений значительно влияют субъективные факторы.

Гистологическая диагностика с помощью интеллектуальных информационно-измерительных систем (ИИС) на базе распознавания образов состоит из нескольких этапов: предобработка изображения объекта, измерение его информативных характеристик, интерпретация результатов (классификация). Предобработка — преобразование входных цифровых изображений, снижающее искажения (помехи,

размытые границы и др.). Объекты измерений в ГД — биопсийный, аутопсийный, операционный, экспериментальный материал и их изображения. В общем случае измеряемые характеристики — отличительные признаки формы, цвета, текстуры, контраста. Интерпретация результатов отражает процесс принятия решений о принадлежности исследуемого объекта к одному из классов заболеваний [1—4]. Результаты измерений интерпретирует врач-патологоанатом или они классифицируются программно с помощью методов сравнения с эталоном, дискриминантных методов, мер близости (сходства) и др. [1, 2, 4].

В ряде практических задач автоматизированной обработки изображений исследуемый объект (группу объектов) предварительно выделяют на фоне общей картины, разбивая изображение на отдельные области, т. е. сегментируя [5]. С методической точки зрения, сегментацию можно рассматривать как составную часть этапа предобработки изображения.

Цель настоящей работы — разработка принципов измерений информативных характеристик объектов и интерпретации результатов с помощью интеллектуальных ИИС ГД онкологических заболеваний.