

УДК 539.1.08; 612.82; 681.322

## ОПРЕДЕЛЕНИЕ ХАРАКТЕРИСТИК ПАМЯТИ

*П.М.Гопыч*<sup>1</sup>

На основе теории проверки статистических гипотез предложены модельное определение и компьютерный метод вычисления характеристик памяти (вероятностей свободного вспоминания, вспоминания с подсказкой, узнавания), определение и способ вычисления интенсивностей подсказок, используемых при тестировании памяти человека. Предложены модели активных и пассивных следов памяти и установлено соотношение между ними. Показано, что ячейку автоассоциативной памяти в виде короткой двухслойной искусственной нейронной сети с повреждениями (или без повреждений) можно использовать для модельного описания характеристик памяти испытуемых с частично поврежденным (или здоровым) мозгом.

### Determination of Memory Performance

*P.M.Gopych*

Within the scope of testing statistical hypotheses theory a model definition and a computer method for model calculation of widely used in neuropsychology human memory performance (free recall, cued recall, and recognition probabilities), a model definition and a computer method for model calculation of intensities of cues used in experiments for testing human memory quality are proposed. Models for active and passive traces of memory and their relations are found. It was shown that autoassociative memory unit in the form of short two-layer artificial neural network with (or without) damages can be used for model description of memory performance in subjects with (or without) local brain lesions.

### 1. ВВЕДЕНИЕ

Важнейшей задачей современной нейробиологии является выявление механизмов кодирования информации в нейронных структурах мозга человека и животных. Из-за междисциплинарности проблемы существенные результаты при ее исследовании часто получают, используя методы смежных дисциплин: эвристического моделирования [1], матричного исчисления [2], термодинамики [3], теории информации [4].

---

<sup>1</sup>Харьковский государственный университет, Украина; e-mail: pmg@pmg.kharkov.com

В настоящей работе на основе статистической теории принятия решений — выбора гипотез (см., например, [5]) предложены модельное определение и способ компьютерного вычисления некоторых количественных характеристик памяти — одной из важнейших когнитивных функций мозга. На важность учета процессов принятия решений при описании функционирования памяти указывалось еще в [6]. Основой предлагаемых результатов стали экспериментальные данные [7] о характеристиках зрительной системы человека-оператора, идентифицирующего (распознающего) одиночные пики на случайном аддитивном фоне в линейчатых спектрах излучения (это один из основных типов первичных экспериментальных данных в современном естествознании), и метод [8] их полного количественного описания с использованием теории искусственных нейронных сетей (см., например, [9, 10]).

Для применения статистической теории принятия решений необходимо в явном виде определить статистические свойства исследуемого объекта. В рамках настоящей работы предполагаем, что сам элемент (ячейка) памяти задается детерминированно, а статистические свойства связанного с ним явления памяти обусловлены статистическими свойствами обрабатываемых им сигналов. Поэтому определим вначале простейшую нейросетевую модель ячейки памяти и модели обслуживающих ее сигналов.

## 2. ПОСТАНОВКА ЗАДАЧИ

Рассмотрим двухслойную искусственную нейронную сеть (ИНС) с одинаковым числом  $N$  входных и выходных нейронов, с синаптическими связями, когда каждый нейрон входного слоя связан со всеми нейронами выходного слоя, со ступенчатой ответной функцией нейронов и нулевым порогом их срабатывания. Полагаем, что модельные нейроны бинарные, т.е. их входные и выходные сигналы принимают только одно из двух дискретных значений:  $-1$  или  $+1$  (можно полагать, что это модельное представление нервных импульсов, поступающих на вход соответственно тормозных и возбуждающих синапсов живых нейронов).

Произвольный набор сигналов на входе (выходе) рассматриваемой ИНС будем записывать в виде векторов  $\mathbf{X}$  в ортогональном пространстве размерности  $N$ , все компоненты (проекции) которых могут принимать значения только  $-1$  или  $+1$ . Тогда совокупность входных сигналов сети — это ее входной вектор  $\mathbf{X}_{in}$ , а совокупность выходных сигналов сети — это ее выходной вектор  $\mathbf{X}_{out}$ . Поскольку  $\mathbf{X}_{in}$ ,  $\mathbf{X}_{out}$  являются бинарными векторами одинаковой размерности  $N$ , то описанная ИНС представляет собой ячейку автоассоциативной [2, 4] памяти емкостью  $N$  бит.

Пусть  $\mathbf{X}_0$  — произвольный вектор, значения проекций которого несут запоминаемую рассматриваемой сетью (ячейкой автоассоциативной памяти) информацию. Будем называть  $\mathbf{X}_0$  вектором-эталоном. Правило, определяющее элементы синаптической матрицы ИНС при запоминании ею эталонного вектора, выберем, следуя Хопфилду [3], в виде

$$w_{ij} = \mathbf{X}_{0i} \mathbf{X}_{0j} / N, \quad (1)$$

где  $\mathbf{X}_{0i}$  —  $i$ -я компонента входного вектора (входной сигнал  $i$ -го нейрона входного слоя сети);  $\mathbf{X}_{0j}$  —  $j$ -я компонента выходного вектора (выходной сигнал  $j$ -го нейрона выходного слоя сети). Таким образом, в (1) предполагаем, что  $\mathbf{X}_{in} = \mathbf{X}_{out} = \mathbf{X}_0$ . По

этой причине  $\mathbf{X}_{0i} = \mathbf{X}_{0j}$  при  $i = j$  ( $i$  — номер входного нейрона,  $j$  — номер выходного нейрона). Видно, что межнейронные связи  $w_{ij}$ , необходимые для запоминания сетью вектора  $\mathbf{X}_0$ , соотношением (1) определяются однозначно, т.е. детерминированно (здесь и далее множитель  $1/N$  несущественен, поэтому в дальнейшем его опускаем).

ИНС с межнейронными связями (1), определяемыми запомненным эталонным вектором  $\mathbf{X}_0$ , будем называть обученной сетью и, следуя [8], будем рассматривать правило (1) вычисления  $w_{ij}$  совместно с определяемыми ниже правилами (2)–(3) преобразования компонент входного вектора как алгоритм распознавания вектора  $\mathbf{X}_0$ . Если после преобразования обученной сетью произвольного входного вектора  $\mathbf{X}_{in}$  на ее выходе возникает вектор  $\mathbf{X}_{out} = \mathbf{X}_0$ , то считаем, что сеть идентифицировала (узнала) запомненный ею эталон  $\mathbf{X}_0$ .

Преобразование входного вектора  $\mathbf{X}_{in}$  в выходной вектор  $\mathbf{X}_{out}$  происходит по таким правилам (см. [3, 8]).

Вычисление входного сигнала для  $j$ -го нейрона выходного слоя сети:

$$H_j = \sum w_{ij} V_i + S_j, \quad (2)$$

где суммирование по  $i = 1, \dots, N$ ;  $V_i$  — выходной сигнал  $i$ -го нейрона входного слоя сети;  $S_j$  — сторонний сигнал (здесь  $S_j = 0$ ).

Вычисление выходного сигнала  $V_j$   $j$ -го нейрона выходного слоя сети ( $j$ -й компоненты выходного вектора  $\mathbf{X}_0$ ):

$$V_j = \begin{cases} +1, & \text{если } H_j > 0 \\ -1, & \text{если } H_j \leq 0. \end{cases} \quad (3)$$

Полагая  $H_i$  равным  $i$ -й компоненте входного вектора  $\mathbf{X}_{in}$ , с помощью (3) находят и выходной сигнал  $V_i$  для  $i$ -го нейрона входного слоя сети.

Предполагаем, что распознаваемый вектор-эталон может быть подвержен случайным искажениям. Тогда для количественного описания этих искажений надо определить понятие шума и правило, определяющее, как этот шум может исказить эталон.

В рамках рассматриваемого подхода проекции всех векторов могут принимать значения только  $-1$  или  $+1$ , поэтому определим случайный вектор, или "шум",  $\mathbf{X}_s$  как вектор, значения компонент которого  $-1$  или  $+1$  распределены случайно и равновероятно. Определим также искаженный эталонный вектор  $\mathbf{X}(d)$  как вектор  $\mathbf{X}_0$ , искаженный (со степенью искажения  $d$ ) шумом  $\mathbf{X}_s$ :

$$\mathbf{X}_i(d) = \begin{cases} \mathbf{X}_{0i}, & \text{если } u_i = 0 \\ \mathbf{X}_{0i}, & \text{если } u_i = 1, \quad \text{sign}(\mathbf{X}_{0i}) = \text{sign}(\mathbf{X}_{si}) \\ \mathbf{X}_{si}, & \text{если } u_i = 1, \quad \text{sign}(\mathbf{X}_{0i}) \neq \text{sign}(\mathbf{X}_{si}), \end{cases} \quad (4)$$

где  $\mathbf{X}_i(d)$ ,  $\mathbf{X}_{0i}$ ,  $\mathbf{X}_{si}$  —  $i$ -е проекции искаженного вектора-эталона, вектора-эталона и шумового вектора;  $i = 1, \dots, N$ ;  $u_i$  — признаки, значения которых "0" или "1" выбираем случайно и равновероятно с соблюдением условия нормировки

$$d = \sum u_i / N. \quad (5)$$

Ясно, что  $0 \leq d \leq 1$ ,  $\mathbf{X}(0) = \mathbf{X}_0$ ,  $\mathbf{X}(1) = \mathbf{X}_s$ .

В силу правила (1) (см. также [3, 8]) при анализе сетью вектора  $\mathbf{X}_0$  на ее выходе обязательно возникает тот же вектор  $\mathbf{X}_0$ , т.е. обученная сеть узнает запомненный ею вектор-эталон с вероятностью 1. При анализе сетью случайного вектора  $\mathbf{X}_s$  (т.е. шума) запомненный ею эталон  $\mathbf{X}_0$  может возникнуть на выходе обученной ИНС только случайно. Следовательно, вероятность  $\alpha$  случайной интерпретации шума в качестве эталона есть  $\alpha < 1$ . При анализе сетью вектора  $\mathbf{X}(d)$  реализуется промежуточная ситуация и на выходе обученной ИНС эталонный вектор  $\mathbf{X}_0$  возникает с вероятностью  $P(d)$ , для которой, по названным причинам, имеем:  $\alpha \leq P(d) \leq 1$ ,  $P(1) = \alpha$ ,  $P(0) = 1$ .

Линейчатые спектры излучения (более общий случай — результат цифрового сканирования произвольных штриховых изображений) можно преобразовать в набор векторов с компонентами из положительных или отрицательных единиц без потери информации, существенной для распознавания в них образов нейросетевыми алгоритмами [8]. При таком преобразовании (по правилам из [8]) участки спектра с хорошо выраженным пиком преобразуются в эталонный вектор  $\mathbf{X}_0$ , участки спектра со слабо выраженным пиком — в искаженный эталонный вектор  $\mathbf{X}(d)$ , со степенью искажения  $d$ , зависящей от отношения сигнал/шум для этого пика, а участки спектра без пиков (фон) — в одну из случайных реализаций вектора  $\mathbf{X}_s$ . Таким образом, определенные выше вектор-эталон, искаженный вектор-эталон и вектор шума действительно реализуются при нейросетевом анализе широко распространенных реальных статистических данных и, более того, все возможные варианты искаженных эталонных векторов  $\mathbf{X}(d)$  исчерпывают все возможные варианты входных векторов  $\mathbf{X}_{in}$  (напомним,  $\mathbf{X}_0$ ,  $\mathbf{X}_s$  являются частными случаями  $\mathbf{X}(d)$ ). Следовательно, если далее вместо  $\mathbf{X}_{in}$  всюду использовать  $\mathbf{X}(d)$ , то общности рассмотрения это не ограничивает.

### 3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Алгоритм распознавания (1)–(3) анализирует входные данные статистического характера, поэтому интерпретируем его как некий статистический критерий проверки простых альтернативных гипотез [5, 8].

Положим, что нулевая гипотеза  $H_0$  состоит в том, что анализируемый сетью входной вектор  $\mathbf{X}_{in} = \mathbf{X}(d)$  есть случайный вектор  $\mathbf{X}_s$ . Тогда альтернативная гипотеза  $H_1$  состоит в том, что  $\mathbf{X}(d)$  есть эталонный вектор  $\mathbf{X}_0$  (хотя и искаженный дискретным бинарным шумом  $\mathbf{X}_s$ ).

Вероятность отвергнуть  $H_0$  есть уровень значимости критерия  $\alpha$  [5], так как согласно приведенному выше определению входной вектор  $\mathbf{X}_s$  интерпретируется сетью как эталонный с вероятностью  $\alpha$ . Случайный вектор  $\mathbf{X}_s$  (т.е. шум) не содержит какой-либо информации об  $\mathbf{X}_0$  и, следовательно,  $\alpha$  есть для обученной сети вероятность "вспомнить" запомненный ею вектор-эталон без подсказки. Роль шумового вектора  $\mathbf{X}_s$  состоит здесь только в том, что он инициирует это вспоминание. Таким образом,  $\alpha$  есть в одно и то же время уровень значимости критерия, вероятность ошибки первого рода, условная вероятность ложного обнаружения (вероятность ложной тревоги) и вероятность вспомнить запомненную информацию без подсказки (вероятность свободного вспоминания из памяти).

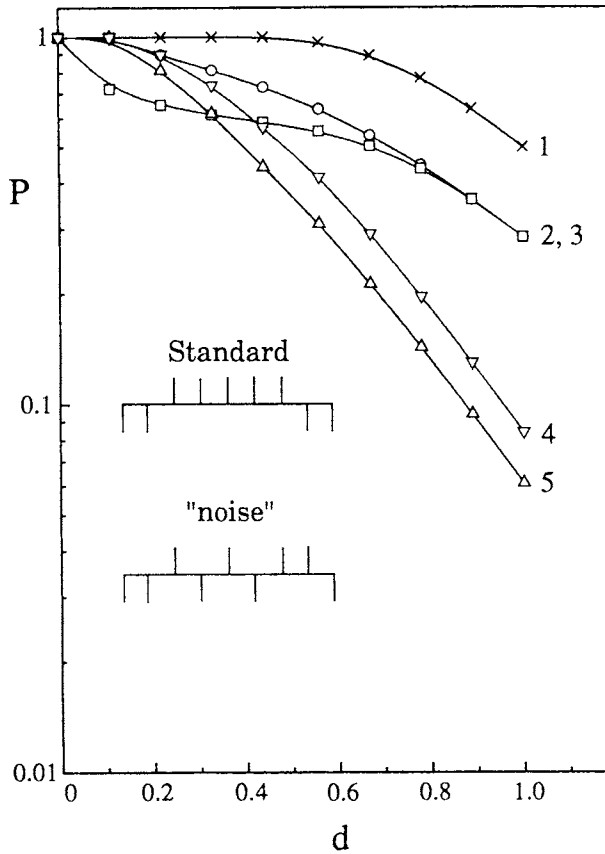
Пусть вероятность отвергнуть  $H_1$ , когда она верна, есть  $\beta$ , т.е.  $\beta$  — это вероятность ошибки второго рода [5]. Вероятность принять гипотезу  $H_1$  при этом же условии

есть мощность критерия  $M = 1 - \beta$ , причем  $M = P(d)$ , так как согласно приведенному выше определению при анализе входного вектора  $\mathbf{X}(d)$  на выходе ИНС эталонный вектор  $\mathbf{X}_0$  возникает с вероятностью  $P(d)$ . Искаженный эталонный вектор  $\mathbf{X}(d)$  содержит некую информацию об  $\mathbf{X}_0$  (тем большую, чем меньше  $d$ ), и эта информация "напоминает" ИНС о запомненном ею эталонном векторе или, другими словами, служит "подсказкой" при вспоминании обученной сетью эталона  $\mathbf{X}_0$ . Следовательно,  $P$  есть вероятность вспоминания сетью эталона  $\mathbf{X}_0$  при наличии подсказки. Интенсивность этой подсказки  $q$  определим как  $q = 1 - d$ . Это простейшее определение  $q$ , для которого выполняется двойное неравенство  $0 \leq q \leq 1$  и  $q(d)$  монотонно растет с уменьшением  $d$  во всем допустимом диапазоне значений  $0 \leq d \leq 1$  (см. выше). Таким образом,  $P(d)$  или  $P(q) = P(1 - d)$  есть в одно и то же время мощность статистического критерия проверки гипотез, условная вероятность истинного обнаружения искаженного вектора-эталона и вероятность вспомнить находящуюся в памяти информацию при степени ее искажения  $d$  или при интенсивности подсказки  $q$ .  $P(0)$  есть при таком рассмотрении вероятность узнавания запомненного в памяти эталона,  $P(1)$  — вероятность его свободного вспоминания (здесь и далее для определенности полагаем  $P = P(d)$ ).

Таким образом, обычно определяемые [11] в нейропсихологических опытах количественные характеристики качества памяти человека — узнавание, свободное вспоминание и вспоминание с подсказкой — приобретают в рамках изложенного подхода строгий количественный смысл. Строгий количественный смысл приобретает и обычно используемое [11] в таких опытах понятие подсказки. Следовательно, для произвольного элемента памяти, независимо от того, принадлежит он живому организму или машине, можно определить ("измерить") характеризующие его качество вероятности  $P(0)$ ,  $P(1)$  и  $P(d)$  при  $0 \leq d \leq 1$ . Для этого надо тестировать память большим количеством запомненных в ней образов с различной степенью их искажения  $d$ . В [7] приведены результаты такого тестирования зрительной системы человека-оператора графическими изображениями коротких фрагментов линейчатых спектров с пиками и без пиков, в [8] приведены результаты такого же тестирования описанных выше ИНС при анализе ими таких же, как в [7], спектров, но в цифровой форме. Методика выполнения упомянутых тестов с помощью компьютерного моделирования подробно описана в [12].

Отметим, что описанная модель памяти предполагает органичное сосуществование двух форм количественного описания следов памяти: 1) в виде синаптической матрицы (1), моделирующей специфическое распределение силы межнейронных связей в области мозга, ответственной за хранение запомненной информации (пассивный след памяти), 2) в виде входных и выходных бинарных векторов, являющихся моделями совокупности нервных импульсов, поступающих на вход или возникающих на выходе той же области мозга (активный след памяти). Матрица синаптических связей (пассивный след памяти) — это модель собственно запоминания (хранения) информации, причем весьма надежного, так как часто даже наличие сильных повреждений обученной сети не исключает полного извлечения запомненной в ней информации (см. ниже). Бинарные векторы (активные следы памяти) моделируют все виды передачи и обработки запомненной информации. Причем тот факт, что эти векторы именно бинарные, означает, что вычисления (обработка информации) в нервных тканях животных и человека выполняются, согласно рассматриваемой модели, в двоичной арифметике.

На рисунке приведены результаты количественного исследования качества описанной выше элементарной ячейки автоассоциативной памяти с помощью векторов  $\mathbf{X}(d)$ , гене-



Вероятности  $P(d)$  вспоминания (узнавания при  $d = 0$ ) сетью с  $N = 9$  запомненного ею эталонного вектора со степенью его искажения  $d$ ; кривые — результат интерполяции вычисленных значений (различные значки). Слева схематически показаны эталонный (Standard) и пример случайного ("noise") векторов: черточки вверх и вниз означают соответственно компоненты этих векторов со значениями +1 и -1. Другие пояснения в тексте

рируемых по правилам (4), (5) при определенных (известных) значениях  $d$  для примера сети с  $N = 9$ . Такое  $N$  выбрано, чтобы облегчить сравнение приводимых и предшествующих [8] результатов и продемонстрировать, что даже при столь малом числе ассоциативных нейронов описанную ИНС можно использовать для количественного моделирования функциональных свойств живого мозга (в отличие от этого в [4] утверждается, что такое моделирование возможно лишь с использованием двухслойных нейронных сетей, содержащих не менее  $10^6$ – $10^7$  ассоциативных элементов с числом синапсов на каждом из них  $\sim 10^5$ – $10^6$ ). На рисунке показаны зависимости  $P(d)$  для неповрежденной (кривая 1) и поврежденных (кривые 2–5) ИНС с  $N = 9$ . Здесь  $P(d) = n_0(d)/n(d)$ , где  $n(d)$  — число разных генерируемых компьютером входных векторов  $\mathbf{X}(d)$  со степенью искажения  $d$  вектора-эталона,  $n_0(d)$  — число случаев, когда в результате анализа сетью входного вектора  $\mathbf{X}(d)$  на ее выходе возникает эталонный вектор  $\mathbf{X}_0$ . Эталонный вектор  $\mathbf{X}_0$  и пример

случайного вектора  $X_s$ , показаны на рисунке внизу слева. При выборе вида  $X_0$  следовали работе [8]. Вероятности  $P(d)$  вычисляли, анализируя все возможные варианты входных векторов  $X(d)$ , так как при  $N = 9$   $n(d)$  может быть полным числом таких вариантов (оно невелико):  $n(d) = 2^m C_N^m$ , где  $d = m/N$ ;  $m \leq N$  — число значений  $u_i = 1$  в формуле (5). Так получаем точные значения  $P(d)$ ; при больших  $N$  можно использовать метод многократных статистических испытаний, но тогда результат вычисления этих вероятностей (частот) будет приближенным. Кривая 1 —  $P(d)$  для ИНС (ячейки памяти) без повреждений. Кривые 2, 3 — варианты  $P(d)$  для ИНС с по-разному случайно выбранными четырьмя ( $\sim 44\%$  от их общего числа  $N = 9$ ) входными нейронами, которые погибли со всеми своими межнейронными связями (в (1) соответствующие  $w_{ij} = 0$ ). Кривые 4, 5 — варианты  $P(d)$  для ИНС без погибших входных нейронов, но с 25 ( $\sim 31\%$  от их общего числа  $N^2 = 81$ ) по-разному случайно выбранными разорванными межнейронными связями.

Из рисунка видно, что для сети без повреждений  $P(1) = 0,5$ , а для сетей с повреждениями  $P(1) < 0,5$ , но тем не менее  $P(1) > 0$ . То есть свободное (без подсказки) вспоминание любой из описанных сетей запомненной в ней информации осуществляется с вероятностью, которая меньше единицы, и поэтому даже при самых благоприятных условиях такого вспоминания (сеть без повреждений) его исход не обязательно положительный. В то же время наличие повреждений обученной сети не исключает возможности безошибочного извлечения запомненной в ней информации, а только делает эту возможность менее вероятной. Так, в частности, проявляется в предлагаемой модели существенно статистический характер процесса извлечения информации из памяти.

Из рисунка также следует, что для неповрежденной сети (кривая 1) и для сетей с повреждениями (кривые 2–5) вероятности  $P(0)$  узнавания запомненной информации одинаковы и, более того,  $P(0) = 1$ . При этом вероятности свободного вспоминания  $P(1)$  и вспоминания с подсказкой  $P(d)$ ,  $0 < d < 1$ , для сетей с повреждениями могут быть существенно меньше, чем для неповрежденной сети. В то же время по результатам систематических экспериментальных исследований установлено [11], что по сравнению с нормальными контрольными испытуемыми у пациентов с частичным повреждением фронтальной доли мозга в наибольшей степени ослабевает способность к свободному вспоминанию стимулов и в наименьшей — способность их узнавания. Эта особенность [11] экспериментальных данных, полученных при исследовании памяти здоровых людей и больных с частичным повреждением фронтальных долей мозга, полностью соответствует указанным выше свойствам значений  $P(0)$  и  $P(1)$  рассматриваемых сетей без повреждений и с повреждениями. Следовательно, такие сети можно использовать для получения количественных результатов при моделировании свойств памяти здорового и частично поврежденного мозга, например, в связи с травмой, болезнью и, возможно, возрастом (предварительные результаты основанного на таком подходе модельного описания свойств удаленной памяти пациентов с частичным повреждением фронтальных долей мозга приведены в [13]).

Описанные ИНС с эталонными векторами, пример которых показан на рисунке (см. [8]), являются также основным элементом компьютерной модели мозга человека, распознающего простые зрительные образы в условиях статистической неопределенности (первое сообщение в [14]).

Предложенный метод компьютерного исследования характеристик ИНС как алгоритмов распознавания (пример результатов для двухслойных сетей показан на рисунке)

можно использовать для выполнения аналогичных исследований элементов памяти любой природы, в частности для исследования сложных ИНС со скрытыми слоями и разным числом нейронов во входном и выходном слое.

В заключение выражаю благодарность академику Н.М.Амосову за внимание к полученным результатам.

### Литература

1. Амосов Н.М. — Алгоритмы разума. Киев: Наукова думка, 1979.
2. Кохонен Т. — Ассоциативная память. М.: Мир, 1980.
3. Hopfield J.J. — In: Proc. Nat. Acad. Sci. USA, 1982, v.79, p.2554.
4. Фролов А.А., Муравьев И.П. — Информационные характеристики нейронных сетей. М.: Наука, 1988.
5. Леман Е.Л. — Проверка статистических гипотез. М.: Наука, 1964.
6. Лурия А.Р. — Нейропсихология памяти. М.: Педагогика, 1974.
7. Гопыч П.М., Сорокин В.И., Сотников В.В. — ПТЭ, 1992, №3, с.95; Instr. Exper. Techn., 1992, v.35, No.3, p.446.
8. Гопыч П.М. — ПТЭ, 1998, №3, с.52; Instr. Exper. Techn., 1998, v.41, No.3, p.341.
9. Кисель И.В., Нескоромный В.Н., Ососков Г.А. — ЭЧАЯ, 1991, т.24, с.1551.
10. Hertz J., Krogh A., Palmer R.G. — Introduction to the Theory of Neural Computation. Redwood City, CA. Addison-Wesley, 1991.
11. Wheeler M.A., Stuss D.T., Tulving E. — In: Journ. of the Int. Neuropsych. Soc., 1995, v.1, p.525.
12. Гопыч П.М. — ЭЧАЯ, 1993, т.24, с.1596; Phys. Part. Nucl., 1993, v.24, p.677.
13. Gopych P.M., Gopych I.P. — In: School of Fund. Med. Journ., Kharkov, 1997, v.3, No.2, p.73.
14. Gopych P.M. — In: School of Fund. Med. Journ., Kharkov, 1996, v.2, No.1, p.91.