МЕХАНИКА
МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

УДК 519.24; 53; 57.017

Искусственные молекулы, собранные из искусственных нейронов, воспроизводящих работу классических статистических критериев

А. И. Иванов[1], А. Г. Банных[2], А. В. Безяев[2]

[1]АО "Пензенский научно-исследовательский электротехнический институт"
Россия, 440000, г. Пенза, ул. Советская, 9
ivan@pniei.penza.ru, (841-2) 59-33-10

[2]Пензенский государственный университет
Россия, 440026, г. Пенза, ул. Красная, 40
ibst@pnzgu.ru; (841-2) 36-82-23; (841-2) 36-84-78

Цель работы показать возможность нейросетевого обобщения множества классических статистических критериев для принятия решений на малых выборках реальных данных. Показано, что для решения задачи необходимо использовать ее предварительную симметризацию, которая позволяет снять проблему моделирования длинных случайных кодов с зависимыми (сцепленными) разрядами. Простота имитационного моделирования симметризованных данных позволяет учитывать корреляционные связи между разрядами случайных кодов и наблюдать ограничения, накладываемые кодами с обнаружением и исправлением ошибок.

Ключевые *слова*: *малые выборки; статистические критерии проверки нормальности данных; сети искусственных нейронов; распознающих нормально распределенные данные.*

DOI: 10.17072/1993-0550-2020-1-26-32

Введение

Если речь идет о натурном эксперименте и получении реальных данных, практически всегда возникает проблема малых выборок. Биолог, подтвердивший свою гипотезу на 16 кроликах (морских свинках, лабораторных крысах или мышах) в глазах своих коллег будет выглядеть не убедительно. На столь малых выборках сегодня нельзя проверить даже гипотезу нормальности распределения данных статистическими критериями, созданными в прошлом веке. По стандартным рекомендациям [1] для доказательной проверки гипотезы нормальности с доверительной вероятностью 0.99 по *хи*-квадрат критерию потребуется выборка в 160 и более опытов.
То же самое относится и к иным не параметрическим статистическим критериям [2].

В качестве иллюстрации проблемы на рис. 1 приведены распределения нормальных и равномерных данных на выходе сумматора искусственного нейрона, воспроизводящего работу *хи*-квадрат критерия.

Рис. 1. *Работа хи-квадрат нейрона, настроенного на разделение малых выборок в 16 опытов нормальных данных и данных с равномерным распределением (доверительная вероятность к решению – 0.66)*

Совершенно иная ситуация сложилась в нейросетевой биометрии.

К результатам работы сети из 256 искусственных нейронов со стороны общества существует высокий уровень доверия. Автоматически обученная по ГОСТ Р 52633.5 [3] сеть искусственных нейронов на 16 примерах образа "Свой" узнает своего хозяина с доверительной вероятностью 0.99 и выявляет попытки предъявления случайных образов "Чужой" с доверительной вероятностью 0.99999. Все это является следствием огромного интереса к биометрии со стороны мирового сообщества, проявляемого в течении последние 30 лет.

Одним из корней доверия к новой технологии является ее высокий уровень стандартизации. Так, по классической статистике в России действует только две рекомендации [1, 2] и один стандарт по терминологии [4], а по биометрии введен в действие 51 стандарт. Формально уровень стандартизации биометрии по сравнению с уровнем стандартизации классической статистики в несколько раз выше.

Все это следствие того, что на биометрию ведущими в информационном отношении государствами (США, Евросоюз, Китай, Россия) за последние 30 лет были затрачены значительные материальные ресурсы. Вполне возможно, что ресурсов, потраченных на решение задач биометрии, за последние 30 лет было больше, чем ресурсов, потраченных мировым сообществом на создание критериев классической статистики за 120 лет ее развития. Пирсон создал *хи*-квадрат критерий в 1900 г., что можно считать началом интенсивного развития классической математической статистики.

Очевидно, что часть наработанных в нейросетевой биометрии и стандартизованных решений можно перенести в классическую статистику. Как показала практика, для каждого из примерно 200 известных статистических критериев [5] может быть построен эквивалентный искусственный нейрон, преобразующий континуум возможных входных состояний малой выборки в один разряд кода (рис. 2).

В итоге мы получаем некоторый статистический аналого-цифровой преобразователь, который можно рассматривать как искусственную статистическую молекулу [6, 7, 8, 9].

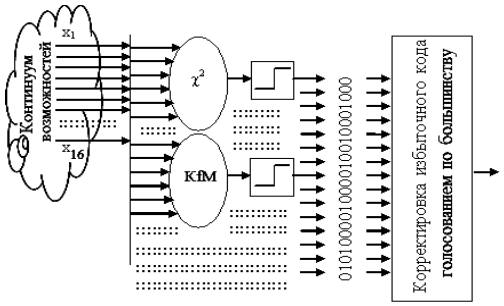

Рис. 2. *Широкая нейронная сеть, состоящая из десятков или сотен нейронов, воспроизводящих работу статистических критериев (нейросетевая математическая молекула)*

Право на такую интерпретацию нейросетевого статистического преобразования дает квантовая химия и квантовая физика, которые рассматривают естественные молекулы, преобразующие континуум собственной энтропии (собственной температуры) в выходные спектральные линии (в конечное число фотонов с дискретными значениями частоты). Нет формальной разницы между дискретными состояниями спектральных линий молекулы водорода и дискретных спектральных линий амплитуд вероятности появления выходных кодовых состояний нейросетевой статистической молекулы рис. 2.

Еще одной аналогией является то, что монография [5] содержит описание примерно 200 известных статистических критериев прошлого века. То есть длина выходного кода искусственной статистической молекулы (рис. 2) уже сегодня может составить порядка 200 бит. Перенос опыта нейросетевого анализа биометрии на решение задач классической статистики показал, что к классическим статистическим критериям могут быть добавлены десятки новых статистических нейросетевых критериев [10, 11, 12]. Можно предположить, что добавление новых статистических критериев позволит в ближайшее время увеличить выходную разрядность искусственных нейросетевых молекул до 256 бит, как это рекомендует пакет из семи национальных стандартов России по нейросетевой биометрии (номера стандартов – ГОСТ Р 52633.xx-20xx). Ориентация на действующие стандарты нейросетевой биометрии при создании нейросетевой статистики искусственных молекул – это попытка сократить достаточно сложный и тернистый путь завоевания доверия к уже отработанной вычислительной технологии в биометрии, но перенесенной в другую предметную область.

Симметризация задачи настройки искусственных нейронов и учета корреляционных связей их выходных состояний

Следует отметить, что наиболее популярные статистические критерии [1, 2], к сожалению, имеют сильные корреляционные связи их решений.

В частности, сильные корреляционные связи имеют следующие статистические критерии:

- *хи*-квадрат критерием (χ^2), $P_1 \approx P_2 \approx P_{EE} \approx 0.344$;
- критерием Крамера-фон Мизеса (KfM), $P_1 \approx P_2 \approx P_{EE} \approx 0.404$;
- критерием Фроцини (Fr), $P_1 \approx P_2 \approx P_{EE} \approx 0.424$;
- критерий Андерсона–Дарлинга (AD), $P_1 \approx P_2 \approx P_{EE} \approx 0.396$;
- логарифмический критерий Андерсона–Дарлинга (ADL), $P_1 \approx P_2 \approx P_{EE} \approx 0.362$.

При оценках работы статистических критериев в группе удобно выполнять симметризацию задачи.

Для этого следует выбирать порог квантования нейронов таким образом, чтобы вероятности ошибок первого и второго рода были близки (почти равновероятны).

Для *хи*-квадрат нейрона квантование выходных данных сумматора нейрона по порогу -64 дает следующее значение вероятностей ошибок:

$$P_1 \approx 0.343 \approx P_2 \approx 0.346 \approx P_{EE} \approx \sqrt{P_1 \cdot P_2} \approx 0.344.$$

Естественно, что каждый критерий (каждый нейрон) будет иметь свое значение порога сравнения и свое значение равновероятных ошибок – P_{EE}. При симметризации данных необходимо вычислять среднее геометрическое всех равновероятных ошибок:

$$\tilde{P}_{EE} = \sqrt[5]{P_{EE}(\chi^2) \cdot P_{EE}(KfM) \cdot P_{EE}(Fr) \cdot P_{EE}(AD) \cdot P_{EE}(ADL)} \approx 0.385.$$

Для учета влияние коэффициентов корреляции данных разных статистических критериев необходимо заменить исходно не симметричную корреляционную матрицу на полностью симметричную:

$$\begin{bmatrix} 1 & r_1 & r_2 & r_3 & r_4 \\ r_1 & 1 & r_5 & r_6 & r_7 \\ r_2 & r_5 & 1 & r_8 & r_9 \\ r_3 & r_6 & r_8 & 1 & r_{10} \\ r_4 & r_7 & r_9 & r_{10} & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & \tilde{r} & \tilde{r} & \tilde{r} & \tilde{r} \\ \tilde{r} & 1 & \tilde{r} & \tilde{r} & \tilde{r} \\ \tilde{r} & \tilde{r} & 1 & \tilde{r} & \tilde{r} \\ \tilde{r} & \tilde{r} & \tilde{r} & 1 & \tilde{r} \\ \tilde{r} & \tilde{r} & \tilde{r} & \tilde{r} & 1 \end{bmatrix},$$

где
$$\tilde{r} \approx E\{|r_i|\} = \frac{2}{n^2 - 2n} \cdot \sum_{i=1}^{0.5n^2 - n} |r_i|. \qquad (1)$$

Данные о коэффициентах корреляции, перечисленных выше классических статистических критериев и о среднем значении их модулей, приведены в табл. 1.

Таблица 1. *Коэффициенты корреляции группы сильно зависимых статистических критериев для малой выборки в 16 опытов*

	χ^2	Fr	KfM	AD	ADL
χ^2	**1**	0.445	0.486	0.392	0.662
Fr	0.445	**1**	0.943	0.613	0.76
KfM	0.486	0.943	**1**	0.666	0.831
AD	0.392	0.613	0.666	**1**	0.698
ADL	0.662	0.76	0.831	0.698	**1**
Усредненные по модулю значения коэффициентов корреляции $\tilde{r} = 0.65$					
Среднее геометрическое вероятностей ошибок $\tilde{P}_{EE} = 0.385$					

Отметим, что сильные корреляционные связи нежелательны (см. табл. 2). В связи с этим можно попытаться выделить группу нейронов (статистических критериев) со слабыми корреляционными связями:

- нейрон Гири (G), $P_1 \approx P_2 \approx P_{EE} \approx 0.231$;
- нейрон *хи*-квадрат (χ^2), $P_1 \approx P_2 \approx P_{EE} \approx 0.344$;
- нейрон Девида–Хартли–Пирсона, (DXP), $P_1 \approx P_2 \approx P_{EE} \approx 0.295$;
- нейрон Лоусена (L), $P_1 \approx P_2 \approx P_{EE} \approx 0.196$;
- нейрон Колмогорова–Смирнова (КС), $P_1 \approx P_2 \approx P_{EE} \approx 0.41$.

Таблица 2. *Коэффициенты корреляции группы слабо зависимых статистических критериев для малой выборки в 16 опытов*

	χ^2	G	DXP	L	KC
χ^2	**1**	-0.016	0.0017	-0.004	0.007
G	-0.016	**1**	0.008	-0.024	0.008
DXP	0.0017	0.008	**1**	0.002	0.015
L	-0.004	-0.024	0.002	**1**	0.008
KC	0.007	0.004	-0.008	0.008	**1**
Усредненные по модулю значения коэффициентов корреляции $\tilde{r} = 0.017$					
Среднее геометрическое вероятностей ошибок $\tilde{P}_{EE} = 0.285$					

Простота имитационного моделирования одинаково коррелированных данных

Из классической теории связи известно, что ошибки в кодах с высокой избыточностью могут быть скорректированы. При этом, чем длиннее код (чем больше его избыточность) тем больше ошибок можно обнаружить и поправить. Нейросетевая молекула рис. 2 способна иметь 256-кратную кодовую избыточность, однако, сколько ошибок в столь длинном коде может быть исправлено – неизвестно.

Все классические коды с обнаружением и исправлением ошибок строились на гипотезе независимости корректируемых разрядов. Мы не можем воспользоваться классикой избыточных кодов, обнаруживающих и исправляющих ошибки. Нам остается только один путь – имитационного моделирования длинных кодов с зависимыми разрядами.

Формально мы можем попытаться решать задачу имитационного моделирования данных с любой асимметричной корреляционной матрицей (1) [13]. Однако это имеет смысл только в том случае, если коэффициенты корреляции заранее вычислены на больших выборках. Если речь идет о вычислении коэффициентов корреляции на малых выборках, задача моделирования становится некорректной, так как вычисление по формуле Пирсона коэффициентов корреляции дает на малых выборках очень большую ошибку. Чем больше размерность задачи – n, тем больше будет ошибка моделирования асимметричных корреляционных матриц.

В этом контексте процедура симметризации задачи через усреднение модулей коэффициентов корреляции (1) является процедурой регуляризации вычислений. То есть, с ростом размерности задачи ошибка имитационного моделирования падает пропорционально $\sqrt{0.5 \cdot n^2 - n}$.

При программной реализации симметричного имитационного моделирования корреляционных связей необходимо воспроизводить множество малых выборок. Функциональные связи имитатора выборок объемом 16 опытов на языке программирования MathCAD отображены на рис. 3.

```
x := | z ← rnorm(16, 0, 1 + rnd(0.01))
     | for i ∈ 0..256
     |   x^⟨i⟩ ← rnorm(16, 0, 1 + rnd(0.01))
     | a ← 1.0
     | for i ∈ 0..256
     |   x^⟨i⟩ ← (z + a·x^⟨i⟩) / sqrt(1 + a^2)
     | x
```

Таблица связи $\tilde{r}(a)$

a	$\hat{r}$
9.37	0.01
2.98	0.10
2.0095	0.20
1.525	0.30
1.222	0.40
1.0	0.50
0.815	0.60
0.654	0.70
0.501	0.80
0.3333	0.90
0.127	0.99

Рис. 3. *Программа, реализующая связывание 256 случайных параметров*

Фрагмент программы связывания 256 векторов данных реализуется с использованием всего одного настраиваемого параметра – "a".

В итоге мы получаем данные с примерно одинаковой взаимной коррелированностью (рис. 4).

x =

	0	1	2	3
0	-0.851	-1.024	0.041	-0.749
1	0.269	-0.447	1.204	-1.662
2	-0.833	-1.14	-2.313	-1.96
3	-0.814	-0.708	-2.629	-0.824
4	-1.076	-0.208	0.716	1.621
5	1.01	-3.085	-2.23	-0.969
6	-1.388	-0.77	-1.054	-1.985

$\text{corr}(x^{\langle 0\rangle}, x^{\langle 1\rangle}) = 0.338$ $\quad \text{corr}(x^{\langle 0\rangle}, x^{\langle 2\rangle}) = 0.532$

$\text{corr}(x^{\langle 1\rangle}, x^{\langle 2\rangle}) = 0.48$ $\quad \text{corr}(x^{\langle 1\rangle}, x^{\langle 3\rangle}) = 0.481$

$\text{corr}(x^{\langle 32\rangle}, x^{\langle 64\rangle}) = 0.211$ $\quad \text{corr}(x^{\langle 64\rangle}, x^{\langle 128\rangle}) = 0.527$

$\text{corr}(x^{\langle 0\rangle}, x^{\langle 256\rangle}) = 0.176$ $\quad \text{corr}(x^{\langle 0\rangle}, x^{\langle 128\rangle}) = 0.338$

$\text{corr}(x^{\langle 256\rangle}, x^{\langle 2\rangle}) = 0.356$ $\quad \text{corr}(x^{\langle 256\rangle}, x^{\langle 37\rangle}) = 0.263$

Рис. 4. *Результаты связывания между собой 256 векторов по 16 случайных данных с нулевым математическим ожиданием и почти единичным стандартным отклонением*

Из рис. 4 видно, что из-за малого объемы выборок (всего 16 отсчетов) корреляционные связи между 256 параметрами кажутся достаточно случайными. На самом деле коэффициенты корреляции одинаковы.

В этом можно убедиться, увеличив примерно в 100 раз объем выборок, на которых вычисляются коэффициенты корреляции по формуле Пирсона. Для этого потребуется внести изменение в программный модуль рис. 3.

Получение длинных кодовых последовательностей квантованием предварительно связанных данных

Исходя из того, что корреляционные связи для 256 нейронов нейросетевой молекулы уже сцеплены программным модулем рис. 3, для получения длинных бинарных последовательностей достаточно выполнить квантование континуальных данных.

Так как данные симметричны, программа квантования (рис. 5) сравнивает данные всех нейронов с одинаковым порогом – "b". Значение порога подбирается таким образом, чтобы среднее значение оцифрованных данных – mean(k) – совпадало со средним геометрическим значением вероятностей ошибок объединяемых нейронов – $\tilde{P}_{EE}$ при многократном запуске программ.

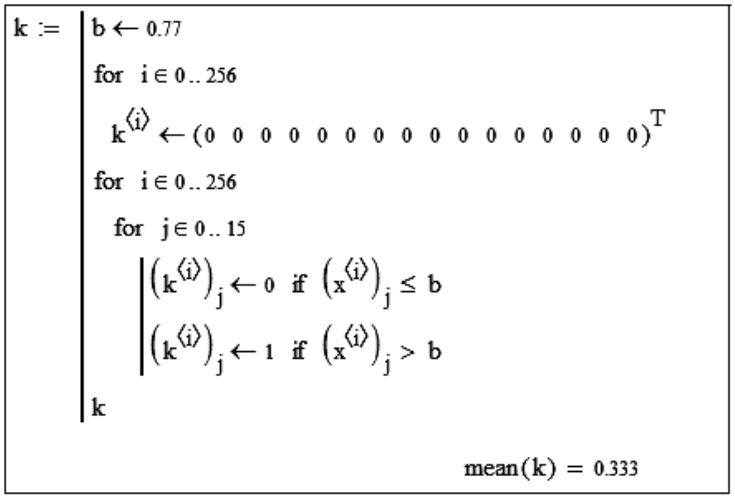

Таблица связей

b	$\tilde{P}_{EE}$
1.6405	0.05
1.27	0.10
1.035	0.15
0.831	0.20
0.6725	0.25
0.525	0.30
0.3825	0.35
0.252	0.45
0.0	0.50

Рис. 5. *Программный модуль квантования данных симметричных нейронов*

В итоге мы получаем длинные кодовые последовательности с коррелированными (сцепленными) между собой разрядами. При этом мы сталкиваемся с проблемой анализа длинных кодовых последовательностей. Если длина последовательностей 256 бит, то мы получаем 2^{256} состояний. Работать со столь длинными кодами сложно, в частности очень сложно вычислить энтропию таких кодов.

Упростить задачу удается, если перейти от самих кодов к расстоянию Хэмминга до интересующего нас идеального кода "0000….0000", соответствующего ситуации, когда все нейроны искусственной молекулы приняли одно и то же решение "0" – нормальная выборка.

Преимуществом перехода от обычных кодов в пространство расстояний Хэмминга является то, что исходное число анализируемых состояний 2^{256} экспоненциально снижается до величины (256+1) состояний.

Программный модуль перехода к расстояниям Хэмминга дан на рис. 6.

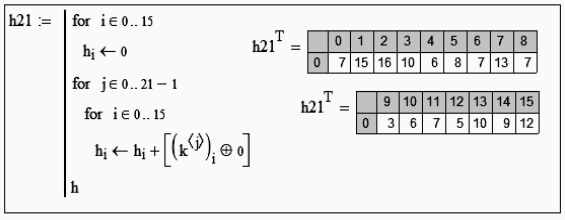

Рис. 6. *Программный модуль вычисления 16 расстояний Хэмминга для 21 нейрона, каждый из которых обучен давать состояние "0" для малой выборки нормальных данных и состояние "1" для выборки равномерных данных*

Использование простейшего кода коррекции ошибок 256-битного выходного кода нейросетевой искусственной молекулы

Следует отметить, что выходные коды нейросетевой молекулы имеют большинство состояний разрядов "0" и примерно 1/3 разрядов "1", если разряды слабо зависимы. Чем больше зависимость разрядов, тем хуже работают классические коды с обнаружением и исправлением ошибок.

В этом контексте искусственную молекулу целесообразно собирать из нейронов со слабо коррелированными откликами. Очевидно, что ориентироваться на худший вариант сильных корреляционных связей классических нейронов (табл. 1) нельзя. Нельзя также ожидать почти полного отсутствия корреляционных связей (табл. 2). Если исходить из гипотезы уровня корреляционных связей – 0.15 и среднего геометрического вероятностей ошибок – 0.33, то распределение расстояний Хэмминга оказывается близко к нормальному распределению (см. рис. 7).

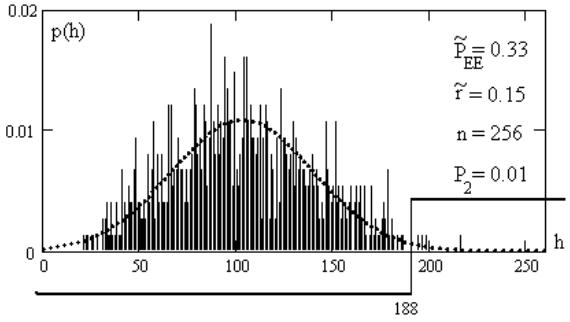

Рис. 7. *Распределение расстояний Хэмминга выходных кодов искусственной статистической молекулы, собранной из 256 искусственных нейронов*

Нормальность распределения данных позволяет легко вычислить порог кода корректировки данных для достижения ошибки второго рода P_2=0.01, что вполне приемлемо для практики. Математическое ожидания расстояний Хэмминга составляет значение 103.3 бита, стандартное отклонение – 36.6 бита. При таком соотношении статистических моментов порог принятия решений составляет 188 бит с состоянием "0". При обнаружении числа "1" более 78 бит (менее 188 состояний "0") принимается решение об отвержении гипотезы о нормальности распределения значений малой выборки в 16 опытов.

Заключение

Таким образом, усложнение вычислений примерно в 256 раз при замене одного *хи*-квадрат критерия на 256 похожих слабо коррелированных преобразований должно позволить снизить вероятность ошибок с 0.344 до величины 0.01, в 34 раза меньше и вполне приемлемо для практики.

В прошлом веке было создано порядка 200 статистических критериев, при этом математики, их создававшие, стремились к повышению мощности каждого из критериев.

В XXI в. появилась возможность нейросетевого объединения множества критериев, однако при этом придется при создании новых статистических критериев учитывать их коррелированность с уже известными критериями.

Кроме того, ожидается появление еще одного дополнительного фактора конкуренции между коммерческими фирмами, такими как: MathCAD, STATISTICA, Maple, MatLAB. Скорее всего, именно фирмы коммерческой математики, конкурируя между собой, решат задачу нейросетевого обобщения множества статистических критериев.

Список литературы

1. *Р* 50.1.037-2002. Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Ч. I. Кри-терии типа χ2. Госстандарт России. М., 2001. 140 с.
2. *Р* 50.1.037-2002. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Ч. II. Непараметрические критерии. Госстандарт России. М., 2002. 123 с.
3. *ГОСТ Р* 52633.5-2011. "Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа".
4. *ГОСТ Р* 50779.10-2000. "Статистические методы. Вероятность и основы статистики. Термины и определения".
5. *Кобзарь А.И.* Прикладная математическая статистика / для инж. и науч. работников. М.: ФИЗМАТЛИТ, 2006. 816 с.
6. *Ахметов Б.Б., Иванов А.И., Фунтикова Ю.В.* Дискретный характер закона распределения *хи*-квадрат критерия для малых тестовых выборок // Вестник Национальной академии наук Республики Казахстан. 2015. № 1. С. 17–25. URL: http://nauka-nanrk.kz/ru/assets/-журнал%202015%201/Вестник_01_2015.pdf (дата обращения: 22.12.2019).
7. *Кулагин В.П., Ахметов Б.Б., Иванов А.И., Газин А.И.* Циклические континуально-квантовые вычисления: усиление мощности *хи*-квадрат критерия на малых выборках // Аналитика. 2016. Т. 30, № 5. С. 22–29. URL: http://www.j-analytics.ru/journal/article/5679 (дата обращения: 22.12.2019).

8. *Иванов А.И., Куприянов Е.Н., Туреев С.В.* Нейросетевое обобщение классических статистических критериев для обработки малых выборок биометрических данных // Надежность. 2019. № 2. С. 22–27. DOI: https: // doi.org/10.21683/1729-2646-2019-19-2-22-27.
9. *Волчихин В.И., Иванов А.И., Безяев А.В., Куприянов Е.Н.* Нейросетевой анализ малых выборок биометрических данных с использованием *хи*-квадрат критерия и критериев Андерсона–Дарлинга // Инженерные технологии и системы. 2019. Т. 29, № 2, С 205–217. DOI: https: //doi.org/ 10.15507/2658-4123.029/2019.02.205-217.
10. *Иванов А.И., Малыгина Е.А., Перфилов К.А., Вятчанин С.Е.* Сравнение мощности критерия среднего геометрического и Крамера-фон Мизеса на малых выборках биометрических данных // Модели, системы, сети в экономике, технике, природе и обществе. 2016. № 2. С 155–158.
11. *Иванов А.И.***,** *Банных А.Г.***,** *Куприянов Е.Н., Лукин В.С., Перфилов К.А., Савинов К.Н.* Коллекция искусственных нейронов эквивалентных статистическим критериям для их совместного применения при проверке гипотезы нормальности малых выборок биометрических данных: сб. науч. стат. по матер. I Всерос. науч.-техн. конф. "Безопасность информационных технологий". 24 апреля. Пенза. 2019. С. 156–164.
12. *Иванов А.И., Перфилов К.А, Лукин В.С.* Нейросетевое обобщение семейства статистических критериев среднего геометрического и среднего гармонического для прецизионного анализа малых выборок биометрических данныхL **с**б. науч. ст. Всерос. науч.-техн. конф. "Информационно-управляющие телекоммуникационные системы, средства поражения и их техническое обеспечение" / под общ. ред. В.С. Безяева. Пенза: АО НПП "Рубин". 2019. С. 50–63.
13. *Шалыгин А.С., Палагин Ю.И.* Прикладные методы статистического моделирования. Л.: Машиностроение, 1986. 320 с.

Artificial molecules assembled from artificial neurons that reproduce the work of classical statistical criteria

A. I. Ivanov[1], A. G. Bannykh[2], A. V. Bezyaev[2]
[1]JSC "PNIEI"; 9, Sovetskaya st., Penza, 440000, Russia
ivan@pniei.penza.ru; (841-2) 59-33-10

[2]Penza State University; 40, Red st., Penza, 440026, Russia
ibst@pnzgu.ru; (841-2) 36-82-23; (841-2) 36-84-78

The purpose of the work is to show the possibility of a neural network generalization of many classical statistical criteria for decision making on small samples of real data. It is shown that to solve the problem it is necessary to use its preliminary symmetrization, which allows you to remove the problem of modeling long random codes with dependent (linked) bits. The simplicity of simulated symmetrized data makes it possible to take into account the correlation between bits of random codes and observe the restrictions imposed by codes with the detection and correction of errors.

Keywords: *small samples; statistical criteria for checking data normality; networks of artificial neurons that recognize normally distributed data.*