

УДК: 004.8, 004.94

## Алгоритм выбора структурных параметров искусственной нейронной сети и объема обучающей выборки при аппроксимации поведения динамического объекта

А. Г. Шумихин, А. С. Бояршинова<sup>а</sup>

Пермский национальный исследовательский политехнический университет,  
химико-технологический факультет,  
Россия, 614000, г. Пермь, Комсомольский пр-т, д. 29  
E-mail: <sup>а</sup> boyarshinovaann@gmail.com

*Получено 03 февраля 2015 г.*

В статье сформулирован обобщенный подход к выбору значений структурных параметров искусственной нейронной сети (ИНС) и объема обучающей выборки, основанный на принципе минимизации количества элементов структуры ИНС и объема обучающей выборки при ограничении на значение показателя качества работы нейросетевой модели динамики объекта. Реализован алгоритм выбора структурных параметров ИНС и построения нейросетевой модели.

Проведена серия вычислительных экспериментов, демонстрирующая применимость алгоритма для построения моделей динамических объектов, в основе которых лежит нелинейная автокорреляционная нейронная сеть.

Ключевые слова: модель динамического объекта, обучающая выборка, искусственная нейронная сеть, топология, обучение, оптимизация структуры искусственной нейронной сети

### Algorithm of artificial neural network architecture and training set size configuration within approximation of dynamic object behavior

A. G. Shumixin<sup>1</sup>, A. S. Boyarshinova<sup>1</sup>

<sup>1</sup>Perm National Research Polytechnic University, 29, Komsomolsky pr., Perm, 614000, Russia

**Abstract.** — The article presents an approach to configuration of an artificial neural network architecture and a training set size. Configuration is based on parameter minimization with constraints specifying neural network model quality criteria. The algorithm of artificial neural network architecture and training set size configuration is applied to dynamic object artificial neural network approximation.

Series of computational experiments were performed. The method is applicable to construction of dynamic object models based on non-linear autocorrelation neural networks.

Keywords: dynamic object model, training set, artificial neural network, architecture, training, optimization of artificial neural network architecture

Citation: *Computer Research and Modeling*, 2015, vol. 7, no. 2, pp. 243–251 (Russian).

## Введение

Моделирование как средство исследования, разработки методов управления технологическими объектами и проектирования автоматизированных систем нашло в настоящее время широкое применение [Соловьев, Литовка, 2013; Лотоцкий, 1991]. Моделирование динамических систем, как правило, основано на знании структуры и характера связей параметров этих систем, которые зачастую многочисленны, что значительно осложняет процесс моделирования. Кроме того, характеристики систем подвержены постепенным изменениям, что требует периодической коррекции моделей [Гартман, Клушин, 2006]. Искусственные нейронные сети являются одним из наиболее эффективных инструментов аппроксимации поведения сложных динамических объектов. ИНС позволяют описать поведение объекта, используя информацию о входных и выходных переменных объекта [Шумихин, Бояршинова, Орехов, 2012]. В общем виде ИНС представляет собой совокупность определенным образом связанных между собой формальных нейронов, состоящих из взвешенного сумматора и функции активации. Структура сети и вид функции активации определяют ее назначение. На рисунке 1 представлена структура нелинейной автокорреляционной нейронной сети, аппроксимирующей поведение нелинейных динамических объектов.

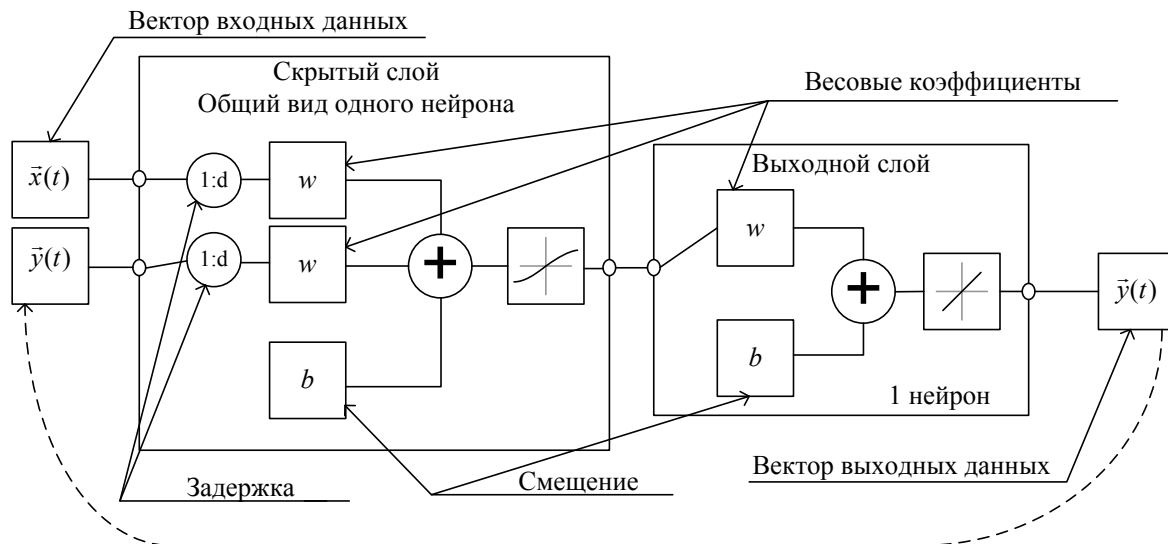


Рис. 1. Структура сети

Сеть является рекуррентной, т. е. принадлежит к классу нейронных сетей, имеющих в своем составе элементы в виде обратных связей и позволяющих моделировать поведение динамических объектов [Макаров и др., 2009]. Временные задержки входных сигналов и сигналов обратных связей позволяют моделировать поведение инерционных объектов с чистым запаздыванием.

В общем виде процесс создания искусственной нейронной сети, представленный на рисунке 2, носит итеративный характер.

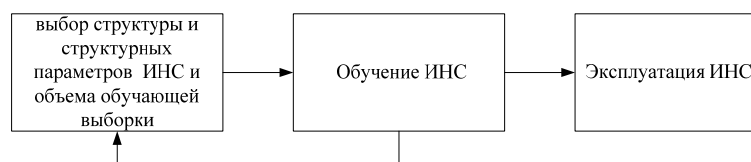


Рис. 2. Процесс создания и использования ИНС

Множество работ посвящено методам обучения нейронных сетей «с учителем» — поиску оптимальных значений синаптических коэффициентов нейронов, осуществляемому на втором этапе создания ИНС [Сараев, 2012].

Первому этапу — выбору структуры ИНС, — напротив, уделено мало внимания. Структурные параметры ИНС и достаточная информативность обучающей выборки во многом определяют работоспособность и эффективность ИНС. Поиск параметров ИНС является итерационной задачей. Зачастую разработчики ИНС подбирают оптимальные значения структурных параметров сети и объем обучающей выборки, основываясь на личном опыте и многократных пробах и ошибках. Поэтому структурные параметры сети и объем обучающей выборки могут оказаться неоптимальными в смысле некоторой функции ошибки аппроксимации.

В ряде работ приведены рекомендации к выбору количества нейронов скрытого слоя в зависимости от размерности задачи [Диагностика информационных..., 2006]. Как правило, эти рекомендации применимы для конкретных рассматриваемых случаев определенной топологии сети.

По формуле, являющейся следствием теоремы Арнольда–Колмогорова–Хехт–Нильсена, можно, например, по известным значениям размерности векторов входных и выходных данных и объему обучающей выборки определить максимальное и минимальное количество необходимых синаптических связей [Круглов, Борисов, 2001]. Формула может быть использована для определения начальных приближений значений структурных параметров нейросетевой модели. Объем обучающей выборки также является параметром, оптимизирующим качество нейросетевой модели.

Чем больше объем обучающей выборки, тем необходимы большие объем памяти для ее хранения, затраты времени на бучение ИНС и сбор информации об объекте. Обучающая выборка с небольшим количеством представителей не обладает достаточной информативностью для того, чтобы с приемлемым качеством характеризовать поведение объекта. Это ведет к тому, что сеть часто оказывается неспособной предсказать поведение объекта вне примеров обучающей выборки.

Недостаточное количество нейронов в скрытом слое не позволит в полной мере аппроксимировать поведение объекта, и ошибка прогнозирования будет велика. Но чем сложнее нейронная сеть, тем больше времени требуется на ее обучение и для работы при решении задачи. Прогнозирующая способность сети может снижаться и в связи с эффектом переобучения. ИНС будет отражать незначительные или несущественные детали в изучаемой зависимости, например шум, выходной вектор будет существенно изменяться при малых отклонениях входного вектора и ИНС будет не способна к обобщению — прогнозированию выходного вектора при векторе входных данных, не вошедших в обучающую выборку.

Таким образом, задача формализации процесса выбора структурных параметров ИНС и объема обучающей выборки представляется актуальной.

В статье сформулирован подход к выбору значений структурных параметров ИНС и объема обучающей выборки, основанный на принципе их минимизации при ограничении на значение показателя качества работы нейросетевой модели.

## Алгоритм поиска параметров сети

При создании нейросетевой модели со структурой, представленной на рисунке 1, необходимо определить количество нейронов во внутреннем слое сети, объем обучающей выборки и значения весовых коэффициентов.

В общем виде блок-схема алгоритма поиска структурных параметров сети и объема обучающей выборки представлена на рисунке 3.

Входными данными для алгоритма являются: минимально допустимое значение коэффициента корреляции между значениями выхода в примерах обучающей выборки и прогнозируемых нейронной сетью, допустимое максимальное значение функции ошибки, начальное значение количества нейронов в скрытом слое и начальное значение объема обучающей выборки.

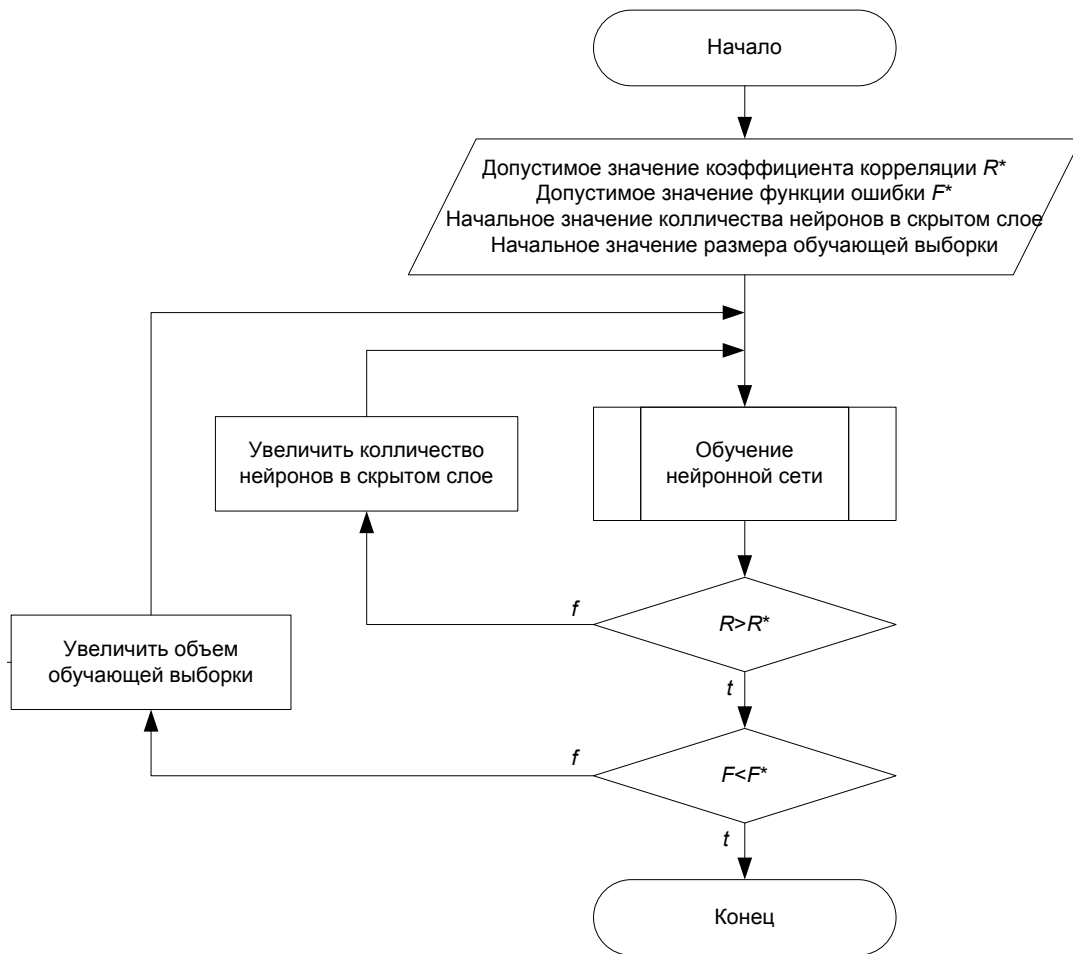


Рис. 3. Блок-схема алгоритма поиска структурных параметров сети и объема обучающей выборки

Процесс обучения нейронной сети по выборке определенного объема для заданного количества нейронов во внутреннем слое является итерационной процедурой. При обучении сети итерационно подбираются значения синаптических весовых коэффициентов до стабилизации значений квадратичной функции ошибки:

$$F_{st}^{er} = F_{st}^{er}(W, W_0, W_1 \dots W_d, b_0, b_1) \approx \text{const}, \quad (1)$$

где  $W, W_0, W_1 \dots W_d$  — массивы весовых коэффициентов выходного слоя, скрытого слоя для вектора входного сигнала и вектора выходного сигнала соответственно;  $b_0, b_1$  — вектор смещений нейронов скрытого и выходного слоя соответственно.

Во внутреннем цикле количество нейронов в скрытом слое увеличивается постепенно до достижения коэффициентом корреляции между значениями выходных данных обучающей выборки с объекта и рассчитанными нейронной сетью некоторого заданного, достаточно высокого, значения. При этом коэффициент корреляции не должен быть слишком высоким. Значение коэффициента корреляции, близкое к единице, скорее указывает на эффект переобучения, когда ИНС отражает практически все, даже незначительные или несущественные, детали в изучаемой зависимости, например шум, но оказывается неспособной к обобщению — прогнозированию выходного вектора при векторе входных данных, не вошедших в обучающую выборку.

Во внешнем цикле объем обучающей выборки увеличивается постепенно до достижения функцией ошибки между выходными значениями тестовой выборки с объекта и рассчитанной сетью некоторого допустимого малого значения.

Таким образом, конструирование сети осуществляется до достижения заданного по значению функции ошибки качества обучения при минимальном количестве нейронов во внутреннем слое и объеме обучающей выборки.

$$F^{er} = F^{er}(W, W_0, W_1 \dots W_d, b_0, b_1, n, m) = F^{er}(F_{st}^{er}(W, W_0, W_1 \dots W_d, b_0, b_1, n, m)) \leq F_{зд}^{er}, \quad (2)$$

где  $F_{зд}^{er}$  — заданное значение функции ошибки.

### Вычислительный эксперимент

С целью иллюстрации изложенного подхода к построению нейросетевой модели, подтверждения работоспособности алгоритма и его практической применимости проведен вычислительный эксперимент на имитационной модели объекта. Модель, реализованная в пакете MATLAB, представлена на рисунке 4.

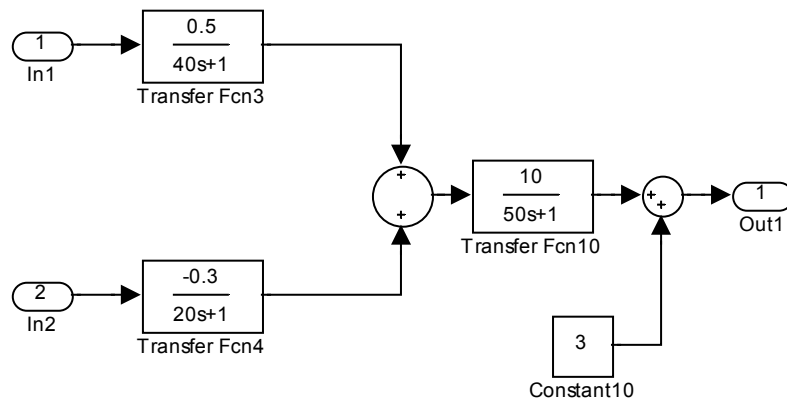


Рис. 4. Имитационная модель динамического объекта

В качестве нейросетевой модели объекта применена нелинейная автокорреляционная нейронная сеть (рисунок 1) с внешними входами, которая является рекуррентной сетью с обратной связью и скрытым слоем, реализующая разностную формулу

$$y[i\Delta t] = F(y[(i-1)\Delta t], y[(i-2)\Delta t], \dots, y[(i-d)\Delta t], x[(i-1)\Delta t], x[(i-2)\Delta t], \dots, x[(i-d)\Delta t]), \quad (3)$$

где  $m = 0, 1, 2, \dots, d$  — такты срабатывания (вычисления) нейронной сети (в эксперименте  $d = 3$ ),  $\Delta t$  — период квантования времени (интервал дискретизации).

Нейронная сеть аппроксимирует поведение объекта, работающего в динамических режимах. Сеть имеет два внешних входа, на которые подаются значения входных сигналов. На вход сети подается также значение выходного сигнала. В сети установлены задержки на 3 такта, что обеспечивает влияние «исторических» данных с глубиной в 3 такта на значение выходной величины объекта. Функцией активации нейронов в скрытом слое является сигмоидальная функция активации — гиперболический тангенс

$$f(n) = \frac{2}{1 + e^{(-2n)}} - 1. \quad (4)$$

Сеть обучалась по методу оптимизации Левенберга–Марквардта. Целевая функция имеет вид

$$F_{st}^{er} = \sum_{i=1}^m (y^*(x_i) - y_i)^2, \quad (5)$$

где  $x_i, y_i$  — значения соответственно входного и выходного сигналов обучающей выборки;  $y^*(x_i)$  — значения выходного сигнала на выходе нейронной сети, рассчитанные по формуле

$$y^*(x_i) = W\varphi\left(f\left(W_0x_i + b_0 + \sum_{k=1}^d(W_k y^*(x_{i-k}) + b_1)\right)\right), \quad (6)$$

где  $W, W_0, W_1 \dots W_d$  — массивы весовых коэффициентов выходного слоя, скрытого слоя для вектора входного сигнала и вектора выходного сигнала соответственно;  $b_0, b_1$  — вектор смещений скрытого слоя и выходного слоя соответственно;  $\varphi, f$  — функции активации выходного и скрытого слоя соответственно;  $d$  — запаздывание.

Выражение для вычисления целевой функции (5), с учетом выражения (6), приобретает вид

$$F_{st}^{er}(W, W_0, W_1 \dots W_d, b_0, b_1) = \sum_{i=1}^m \left( W\varphi\left(f\left(W_0x_i + b_0 + \sum_{k=1}^d(W_k y^*(x_{i-k}) + b_1)\right)\right) - y_i \right)^2. \quad (7)$$

При обучении сети обучающая выборка разбивается случайным образом на три подвыборки (обучающую, контрольную и тестовую) [Круглов, Дли, Голунов, 2000] в соотношении 75 %, 15 % и 15 % соответственно. Начальные значения весовых коэффициентов задаются случайным образом.

Многократное обучение ИНС, вследствие различных начальных значений весовых коэффициентов и различного разбиения обучающей выборки на подвыборки, дает различные результаты. Поэтому на этапе обучения сети (рисунок 3) целесообразно проводить обучение сети несколько раз при одних и тех же параметрах. Необходимо также для каждого цикла работы алгоритма задать максимальное количество итераций с тем, чтобы предотвратить заикливание на случай невозможности построения нейросетевой модели при заданных параметрах. Количество итераций подбирается опытным путем.

Проведена серия из 25 вычислительных экспериментов по построению нейросетевой модели при различных допустимых значениях коэффициента корреляции. Начальные значения количества нейронов в скрытом слое — 5, объема обучающей выборки — 500 точек. Количество нейронов в скрытом слое на шаге поиска увеличивалось на 1 нейрон, объем обучающей выборки увеличивался на 100 точек. Допустимое значение функции ошибки  $F^*$  принято равным 0,05. На этапе обучения, при работе каждого внутреннего цикла, сеть обучалась трижды. Результаты представлены в таблице.

В результате серии вычислительных экспериментов 20 из 25 (80 %) экспериментов оказались успешными — получены нейросетевые модели, аппроксимирующие поведение динамического объекта с ошибкой, не превышающей заданное предельное значение (неудачные эксперименты выделены жирным шрифтом).

Наилучшие нейросетевые модели, в смысле минимизации количества нейронов и объема обучающей выборки, получены при не самых высоких значениях коэффициента корреляции между значениями выхода в примерах обучающей выборки и прогнозируемых нейронной сетью (выделены жирной рамкой). Наихудшие нейросетевые модели, в смысле минимизации количества нейронов и объема обучающей выборки, получены при высоком значении коэффициента корреляции между значениями выхода в примерах обучающей выборки и прогнозируемых нейронной сетью ( $R^* = 0.99$ ).

Из данных таблицы также следует, что у всех выбранных, выделенных контуром, нейросетевых моделей продолжительность работы алгоритма обучения на порядок меньше, чем у остальных, более сложных.

На рисунке 5 представлен процесс изменения показателя качества нейросетевой модели и структурных параметров нейронной сети в ходе первого вычислительного эксперимента. Точки соответствуют значениям функции ошибки между выходными значениями тестовой вы-

борки с объекта и рассчитанными сетью при значениях структурных параметров на текущем этапе вычисления. В результате работы алгоритма за 8 шагов построена нейросетевая модель имитационного объекта, удовлетворяющая заданному показателю качества ( $F^{er} \leq 0.05$ ), параметры которой следующие (1-ая строка таблицы): допустимое значение коэффициента корреляции  $R^* = 0.85$ ; значение функции ошибки  $F^{er} = 0.011$ ; объем обучающей выборки — 1000 представителей; продолжительность работы алгоритма обучения — 228 с.

Таблица. Результаты серии вычислительных экспериментов

Допустимое значение коэффициента корреляции, $R^*$	Значение функции ошибки нейросетевой модели, $F^{er}$	Объем обучающей выборки	Количество нейронов в скрытом слое	Продолжительность работы алгоритма, с
0.85	0.011	1000	7	227.7065
	<b>0.094</b>	<b>4000</b>	<b>9</b>	<b>3194.6067</b>
	0.029	1700	7	569.8345
	0.081	3100	7	1800.156
	0.015	1100	5	164.1851
0.90	0.031	1800	6	549.6854
	0.042	1100	8	295.5246
	0.042	900	5	94.1262
	<b>0.083</b>	<b>3100</b>	<b>7</b>	<b>1649.6087</b>
	0.047	2000	7	673.1184
0.95	<b>0.077</b>	<b>2900</b>	<b>9</b>	<b>1790.0505</b>
	0.034	2600	8	1245.073
	0.026	2400	8	1245.1063
	<b>0.085</b>	<b>2900</b>	<b>9</b>	<b>1738.9558</b>
	0.015	2500	8	1087.7622
0.98	0.049	1700	9	588.8422
	0.027	2000	10	1053.6343
	0.029	900	7	172.1255
	0.023	2300	9	1243.4043
	0.034	2300	9	1200.7219
0.99	0.042	1500	9	672.7662
	<b>0.062</b>	<b>2300</b>	<b>15</b>	<b>1809.1465</b>
	0.013	1900	11	1070.5985
	0.043	1900	12	1043.7941
	0.036	2100	14	1563.0102

## Выводы

Сформулированный обобщенный подход к выбору значений структурных параметров искусственной нейронной сети и объема обучающей выборки основан на принципе минимизации количества элементов структуры искусственной нейронной сети и объема обучающей выборки при ограничении на значение показателя качества работы нейросетевой модели динамики объекта. Реализация данного подхода на примере нелинейной автокорреляционной нейронной сети, аппроксимирующей поведение имитационного динамического объекта в результате серии вычислительных экспериментов показала применимость изложенного подхода при оптимиза-

ции структуры и объема обучающей выборки при построении модели динамических объектов, в основе которых лежит нелинейная автокорреляционная нейронная сеть.

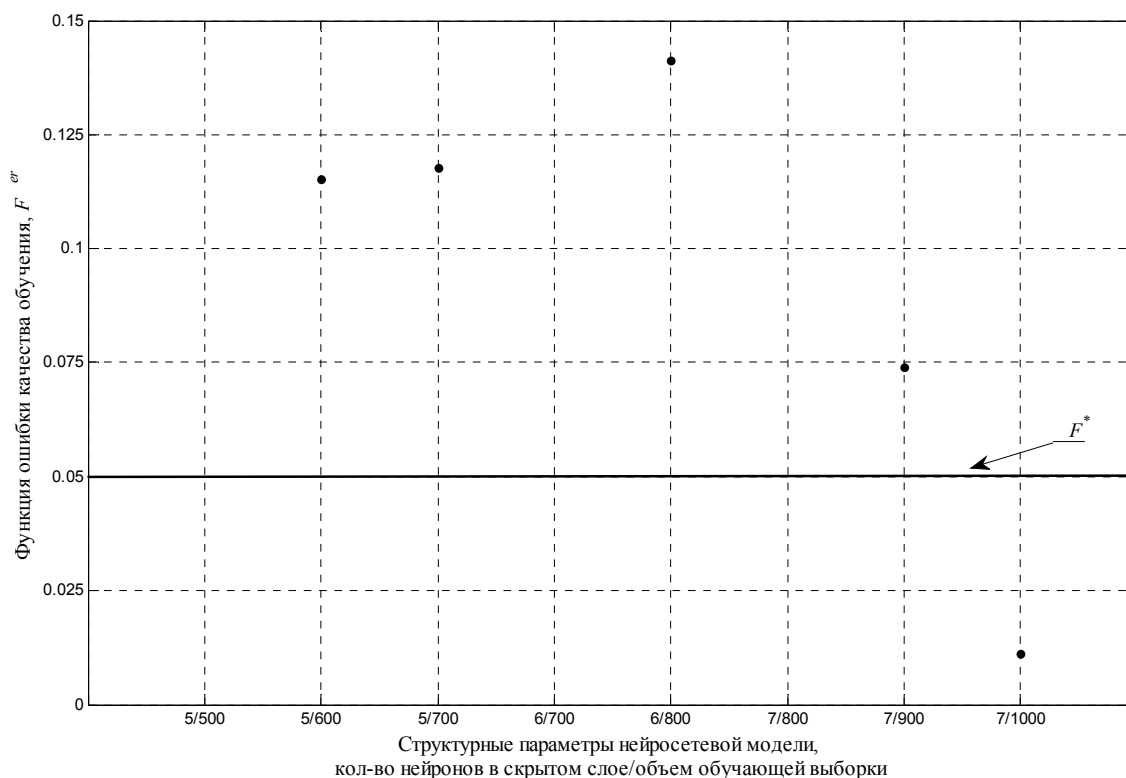


Рис. 5. Изменение показателя качества нейросетевой модели и структурных параметров нейронной сети в процессе вычислительного эксперимента

## Список литературы

- Гартман Т. Н., Клушин Д. В. Основы компьютерного моделирования химико-технологических процессов: учеб. пособие для вузов. — М.: ИКЦ «Академкнига», 2006.
- Круглов В. В., Борисов В. В. Искусственные нейронные сети. Теория и практика. — 1-е. — М.: Горячая линия. — Телеком, 2001. — С. 382.
- Круглов В. В., Дли М. И., Голунов Р. Ю. Нечеткая логика и искусственные нейронные сети. — М.: Физматлит, 2000. — 224 с.
- Лотоцкий В. А. Идентификация структур и параметров систем управления // ИКА. — 1991. — № 3–4. — С. 30–38.
- Макаров И. М., Лохин В. М., Манько С. В., Романов М. П. Автоматизация синтеза и обучение интеллектуальных систем управления / под ред. И. М. Макарова и В. М. Лохина. — М.: Наука, 2009.
- Ретин А. И., Сабанин В. Р., Смирнов Н. И., Андреев С. Н. Диагностика информационных подсистем АСУТП с использованием технологий искусственного интеллекта // Теплоэнергетика. — 2006. — № 6. — М.: ООО МАИК «НАУКА/ИНТЕРПЕРИОДИКА».
- Сараев П. В. Численные методы интервального анализа в обучении нейронных сетей // АиТ. — 2012. — № 11. — С. 129–143; Saraev P. V. Numerical methods of interval analysis in learning neural network, Autom. Remote Control, 73:11, 1865–1876.



*Соловьев Д. С., Литовка Ю. В.* Математическое моделирование и оптимальное управление процессом осаждения гальванического покрытия в многоанодной ванне с учетом изменения концентрации компонентов электролита // Компьютерные исследования и моделирование. — 2013. — Т. 5, № 2. — С. 193–203.

*Шумихин А. Г., Бояришинова А. С., Орехов М. С.* Применение нейросетевых моделей при автоматизированном управлении сложными химико-технологическими системами // Ползуновский вестник. — 2012. — № 3/2. — С. 9–12.