

## ЗАВИСИМОСТЬ ПРОИЗВОДИТЕЛЬНОСТИ НЕЙРОСЕТИ С ПРЯМОЙ СВЯЗЬЮ С ОДНИМ СКРЫТЫМ СЛОЕМ НЕЙРОНОВ ОТ ГЛАДКОСТИ ЕЁ ОБУЧЕНИЯ НА ЗАШУМЛЕННЫХ КОПИЯХ АЛФАВИТА ОБРАЗОВ

Рассматривается нейросеть с прямой связью с одним скрытым слоем нейронов, предназначенная для распознавания монохромных изображений небольшого формата. Исследуется зависимость производительности такой нейросети от гладкости её обучения, определяемой в смысле частоты разбиения отрезка среднего квадратического отклонения шума для генерирования и подачи зашумленных изображений на вход. Полученные по достаточно объёмным тестированиям нейросети результаты свидетельствуют о том, что производительность нейросети при сглаживании процесса обучения повышается, однако при этом процесс обучения значительно замедляется. В средах с более мощными вычислительными ресурсами это замедление предполагается менее ощутимым, поэтому там общая функциональность нейросети при гладком обучении будет улучшена.

*Ключевые слова:* нейросеть, распознавание образов, обучение, гладкое обучение, производительность, скорость обучения.

*There is considered the feed-forward neuronet with single hidden layer, assigned for recognition of monochrome images of a not large format. There is investigated the dependence of such neuronet performance against its training smoothness, defined in the sense of break frequency for noise mean-square deviation segment in generating and feeding the input with noised images. The obtained results after enough-voluminous neuronet testings are evidence that neuronet performance by smoothing the training process improves, but at that the training process slows down conspicuously. Within environments with more powerful computational resources this deceleration is surmised to be less perceptible, and so the general neuronet functionality there by the smooth training will be improved.*

*Keywords:* neural network, pattern recognition, learning, smooth learning, performance, speed of learning.

### Проблема подстройки параметров нейронной сети

Известно, что искусственная нейронная сеть имеет достаточно большое множество регулируемых свойств и макропараметров: тип архитектуры, объёмность скрытого слоя нейронов, количество входов и выходов, специфика метода обучения [1, 2]. Поиск методов оптимальной настройки параметров нейронных сетей на сегодня вырос в очень сложную проблему, напоминающую задачу подбора наилучшего варианта обучения в более глобальном масштабе [3, 4]. Основным критерием оптимального выбора нейросетевого объекта является максимизация его производительности, выражаемая не только в числовых векторах, но и в корректно свёрнутых в одну точку значениях. Однако влияние нескольких макропараметров нейронной сети в совокупности даже на свёрнутое в одну точку значение производительности неизвестно [1], из-за чего приходится изучать подстройку каждого параметра в отдельности.

### Метод подстройки параметра нейронной сети

Естественным методом подстройки параметра искусственной нейронной сети является программирование и циклическая прогонка её функционирования в пределах некоторого интервала или множества значений (свойств или атрибутов) исследуемого параметра [1]. Когда удастся найти то значение (свойство или атрибут) параметра, при котором средняя производительность нейронной сети достигает максимального уровня, тогда это значение считается оптимальным и используется в дальнейшем [1, 3, 5]. Впрочем, не исключено, что подстроенное значение или свойство исследуемого параметра окажется менее эффективным при функционировании нейронной сети с несколько изменёнными входными данными (образами или объектами для исследования и принятия решений). Следовательно, надёжность подстройки параметра искусственной нейронной сети зависит от разнообразия решаемых ею задач [2, 5, 6]. Особенно это ощутимо в процессах обучения, где даже незначительное изменение класса входных данных требует переобучения или дообучения нейросети, поскольку иначе она будет работать непродуктивно. И именно качество процесса обучения определяет основные свойства продуктивности нейросетевого исследования. В частности, для задачи распознавания образов, подаваемых в основном в зашумленном виде, очень важно найти свойство или атрибут процесса обучения, наиболее остро воздействующие на его результаты.

### Постановка задачи

Рассмотрим простую нейронную сеть с прямой связью с одним скрытым слоем нейронов, предназначенную для распознавания монохромных изображений. Общее количество изображений равняется 26. Каждое изображение представляется в форме  $7 \times 5$ -матрицы с элементами 1 и 0. Для решения задачи распознавания спроектируем нейронную сеть с прямой связью и с одним скрытым слоем из 10 нейронов в среде Matlab 7.0.1 с помощью стандартного демо-скрипта "appc1" [7, с. 199 — 206] из инструмента Matlab Neural Network Toolbox. Сеть должна научиться распознавать не только 26 эталонных мини-изображений, но и их зашумленные варианты, где шум отождествляется с естественными изменениями эталонов, приводящих к их отличиям от входных образов. Для этого нейросеть сначала обучается на множестве незашумленных эталонных изображений, а потом к этим эталонам прибавляют гауссов шум [5, 8], и входными образами для обучающейся нейросети уже служат зашумленные 26 монохромных изображений.

Уровень шума лежит в известных пределах, при которых нейросеть ещё способна давать приемлемую производительность. Изменяя его от нулевого значения до предельно максимального с некоторым шагом, нейросеть обучается распознавать расширенный класс монохромных изображений на основе исходных 26. Необходимо выяснить, как этот шаг влияет на качество обучения, то есть, зависит ли и, если зависит, то насколько, производительность нейросети от ступенчатости (гладкости) обучения.

#### Исследование зависимости производительности нейросети от гладкости обучения

Нейросеть должна безошибочно распознавать идеальные векторизированные входные образы и с максимальной точностью воспроизводить зашумленные эталоны. С целью этого исследуемую нейросеть обучают с применением двух идеальных и двух зашумленных копий алфавита из 26 векторизированных входных образов [7, с. 203]. Таким образом, во время обучения на вход нейросети поступает множество

$$P_{\text{train}} = \{\mathbf{P}_i\}_{i=1}^4 = \{\mathbf{A}, \mathbf{A}, \mathbf{A}_1, \mathbf{A}_2\} \quad (1)$$

из  $35 \times 26$ -матрицы  $\mathbf{A}$  всех эталонов, а также матриц

$$\mathbf{A}_1 = \mathbf{A} + \sigma_1 \mathbf{\Xi}, \quad \mathbf{A}_2 = \mathbf{A} + \sigma_2 \mathbf{\Xi}, \quad (2)$$

формируемых при некоторых средних квадратических отклонениях  $0 < \sigma_1 < \sigma_2$  для  $35 \times 26$ -матрицы  $\mathbf{\Xi}$  из значений нормально распределённой случайной величины с нулевым математическим ожиданием и единичной дисперсией. Поскольку предполагается, что среднее квадратическое отклонение шума не превосходит 0.2 [7, с. 200], то естественным образом полагают  $\sigma_2 = 0.2$  и, как следствие элементарного усреднения,  $\sigma_1 = 0.1$ . Во множество целевых матриц (идентификаторов распознавания)

$$T = \{\mathbf{T}_i\}_{i=1}^4 = \{\mathbf{I}, \mathbf{I}, \mathbf{I}, \mathbf{I}\} \quad (3)$$

входит только единичная  $26 \times 26$ -матрица  $\mathbf{I}$ , и множество (3) обеспечивает идентификацию каждого столбика матриц множества (1). Процедура отклика нейросети на входное множество (1) с элементами (2) при идентификаторе (3) повторяется  $K$  раз. В демо-скрипте “arprc1” эмпирически положено  $K = 10$ , чего вполне достаточно для качественной работы нейросети.

Зависимость производительности нейросети от уровня шума  $\sigma > 0$  на её входе

$$P = \{\mathbf{A} + \sigma \mathbf{\Xi}\} \quad (4)$$

понимается как среднестатистический процент ошибок распознавания (СПОР)  $\gamma(\sigma, N)$ , зависящий ещё и от объёма  $N$  тестирования нейросети. При этом полезно сравнивать функцию  $\gamma(\sigma, N)$  с аналогичной функций распознавания  $\gamma_0(\sigma, N)$ , производимой нейросетью, которая была обучена в бесшумных условиях. На рис. 1 — 4 показаны статистические кривые  $\gamma(\sigma, N)$  и  $\gamma_0(\sigma, N)$ , полученные тестированием нейросети, обученной Matlab-методом `traindga` по множеству (1) и с множеством идентификаторов (3). Как видим из рис. 1, 2 и 4, чёткая эффективность обучения нейросети по множеству (1) отслеживается лишь при  $\sigma > 0.2$ , чего, кстати, не предусматривает постановка задачи, так что для неё это преимущество перед бесшумным обучением оказывается невостребованным. А на рис. 2 и 4 изображён статистический случай того, когда при  $\sigma < 0.18$  обучение по множеству (1) вообще неэффективно, поскольку СПОР нейросети, обученной в бесшумных условиях, здесь зафиксирован меньшим.

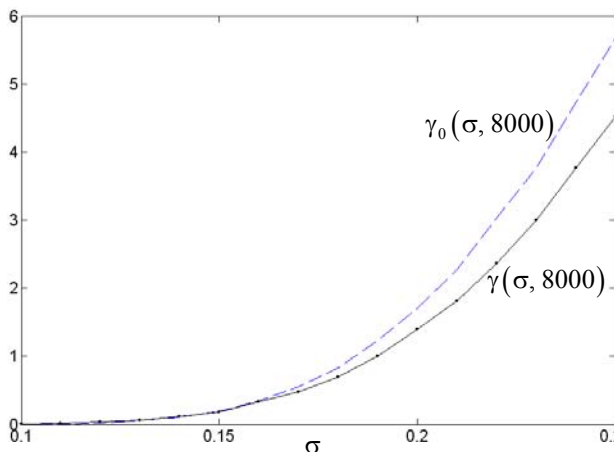


Рис. 1. СПОР нейросети, обученной по множеству (1), с общим временем функционирования 1221.438 с

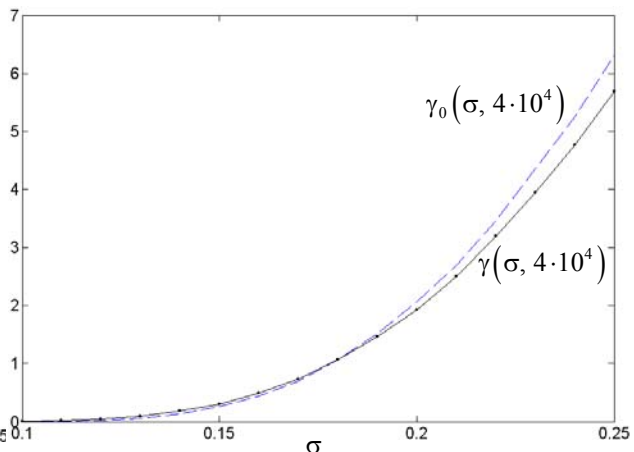


Рис. 2. СПОР нейросети, обученной по множеству (1), с общим временем функционирования 6068.39 с

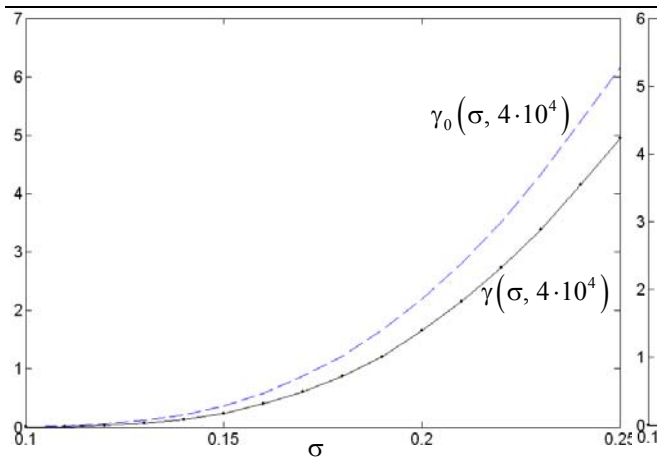


Рис. 3. СПОР нейросети, обученной по множеству (1), с общим временем функционирования 6110.937 с

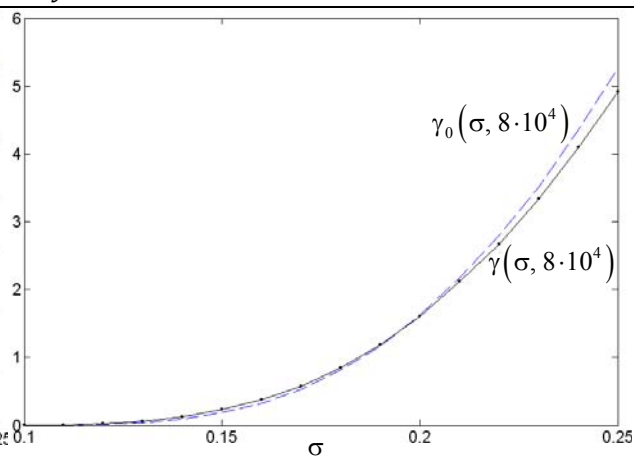


Рис. 4. СПОР нейросети, обученной по множеству (1), с общим временем функционирования 12190.921 с

Статистические кривые  $\gamma(\sigma, N)$  и  $\gamma_0(\sigma, N)$ , полученные тестированием нейросети, обученной Matlab-методом `traingda` по множеству

$$\tilde{P}_{\text{train}} = \{\tilde{\mathbf{P}}_i\}_{i=1}^{10} = \{\mathbf{A}, \mathbf{A}, \{\mathbf{A}_k\}_{k=1}^8\} \quad (5)$$

из  $35 \times 26$ -матрицы  $\mathbf{A}$  всех эталонов и матриц

$$\mathbf{A}_k = \mathbf{A} + \sigma_k \mathbf{\Xi} \quad (6)$$

при

$$\sigma_k = 0.025k \quad \forall k = \overline{1, 8} \quad (7)$$

с множеством

$$T = \{\mathbf{T}_i\}_{i=1}^{10} = \{\mathbf{I}, \mathbf{I}, \mathbf{I}, \mathbf{I}, \mathbf{I}, \mathbf{I}, \mathbf{I}, \mathbf{I}, \mathbf{I}, \mathbf{I}\} \quad (8)$$

идентификаторов распознавания, показаны на рис. 5 — 8. И, как следует из рис. 5 — 7, обучение нейросети по множеству (5) является эффективным для  $\sigma > 0.12$ , но на рис. 8 изображён статистический случай того, что только при  $\sigma > 0.15$  такое обучение предпочтительно. Тем не менее, сравнивая кривые  $\gamma(\sigma, N)$  при  $N \in \{8000, 4 \cdot 10^4, 8 \cdot 10^4\}$  на рис. 1 — 4 и рис. 5 — 8, нетрудно убедиться в том, что обучение нейросети по множеству (5) более эффективно по сравнению с обучением по множеству (1). Это наглядно демонстрирует рис. 9 со всеми статистическими кривыми СПОР  $\gamma(\sigma, N)$  и  $\gamma_0(\sigma, N)$  из рис. 1 — 8 при  $N \in \{8000, 4 \cdot 10^4, 8 \cdot 10^4\}$ . При том время, затраченное на более гладкое обучение по (5) — (8) и тестирование нейросети, вполне сравнимо со временем её жёсткого обучения по (1) — (3) и тестирования, а для самого длинного статистического наблюдения с  $N = 8 \cdot 10^4$  более гладко обученная нейросеть функционировала в общей сложности быстрее на 24 с (рис. 4 и рис. 8).

Несомненно, исследованный пример свидетельствует о том, что производительность нейросети зависит от гладкости её обучения, где под гладкостью следует понимать частоту разбиения предполагаемого отрезка среднего квадратического отклонения шума, в точках которого будет сгенерирован соответствующий шум для входа (4). Но в данном случае обучение нейросети Matlab-методом `traingda` по множеству (5) из (6) при (7) с множеством идентификаторов (8) не обязательно является оптимальным решением или близким к таковому. Вероятно, для каждой задачи распознавания монохромных изображений или, вообще говоря, распознавания образов при предполагаемом максимальном среднем квадратическом отклонении шума  $\sigma_{\text{max}}$  существуют такие  $C \in \mathbb{N}$  и  $F \in \mathbb{N}$ , что обучение нейросети некоторым методом по множеству

$$\tilde{P}_{\text{train}} = \{\tilde{\mathbf{P}}_i\}_{i=1}^{C+F} = \{\{\mathbf{A}\}_{i=1}^C, \{\mathbf{A}_k\}_{k=1}^F\} \quad (9)$$

из  $C$  матриц  $\mathbf{A}$  всех эталонов и матриц

$$\mathbf{A}_k = \mathbf{A} + \sigma_k \Xi$$

зашумлених образів при

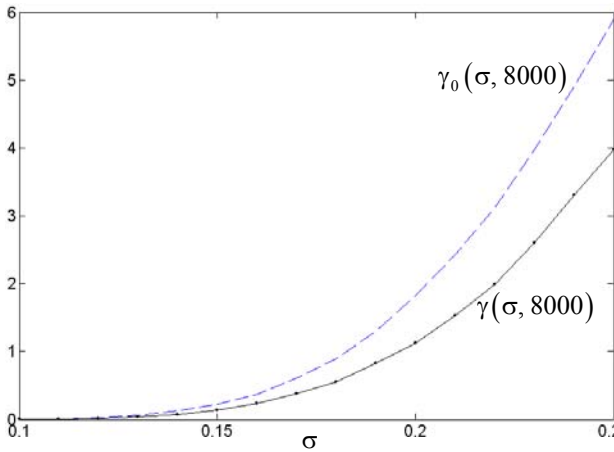


Рис. 5. СПОР нейросети, обученной по множеству (5), с общим временем функционирования 1234.25 с

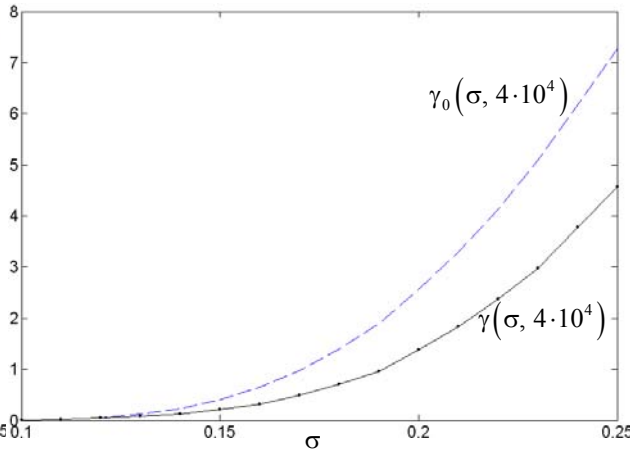


Рис. 6. СПОР нейросети, обученной по множеству (5), с общим временем функционирования 6071.75 с

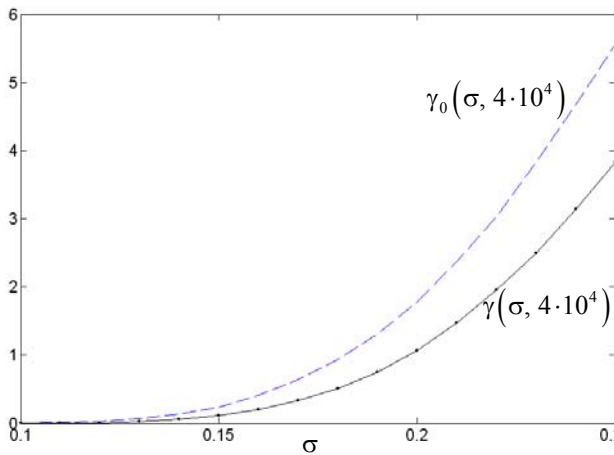


Рис. 7. СПОР нейросети, обученной по множеству (5), с общим временем функционирования 6109.328 с

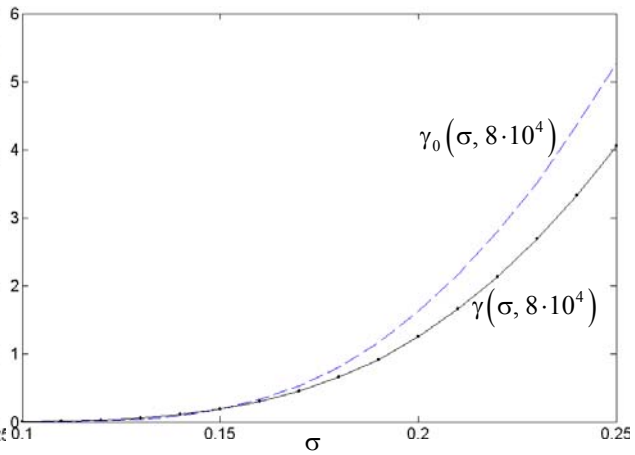


Рис. 8. СПОР нейросети, обученной по множеству (5), с общим временем функционирования 12166.046 с

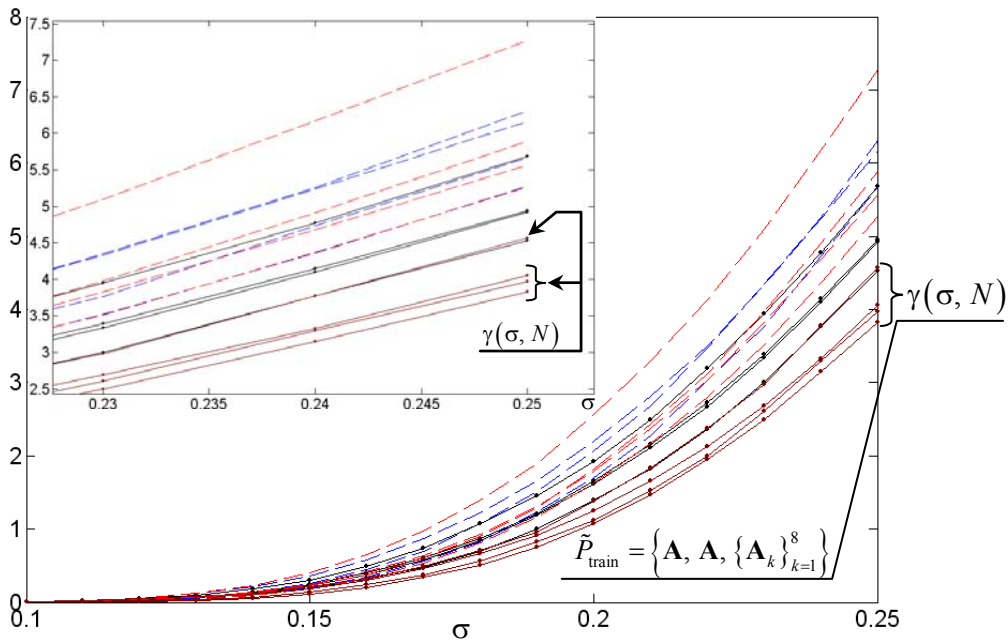


Рис. 9. Все статистические кривые СПОР  $\gamma(\sigma, N)$  и  $\gamma_0(\sigma, N)$  из рис. 1 — 8 при  $N \in \{8000, 4 \cdot 10^4, 8 \cdot 10^4\}$ , полученные жёстким обучением нейросети по (1) — (3) и более гладким обучением нейросети по (5) — (8)

$$\sigma_k = \frac{\sigma_{\max}}{F} \cdot k \quad \forall k = \overline{1, F} \quad (11)$$

с множеством

$$T = \{\mathbf{T}_i\}_{i=1}^{C+F} = \{\mathbf{I}\}_{i=1}^{C+F} \quad (12)$$

идентификаторов минимизирует функцию  $\gamma(\sigma, N) \quad \forall \sigma \in (0; \sigma_{\max}]$  (при достаточно большом  $N$ ). Вопрос о том, насколько при этом затягивается процесс обучения, должен изучаться отдельно.

Результаты тестирования нейросети при повышении  $C$  не дают оснований считать, что СПОР особо чувствителен к количеству эталонных матриц в (9) на входе нейросети (рис. 10). По крайней мере, более определяющим здесь усматривается число  $F$  (рис. 11), где уже при  $F = 288$  максимальный СПОР снижается до 2, что практически вдвое меньше максимального СПОР при  $F = 8$ . Однако рост времени обучения происходит значительно быстрее, чем снижается СПОР (рис. 12), и уже при  $F = 288$  нейросеть обучается почти в 65 раз медленнее, чем при  $F = 8$ .

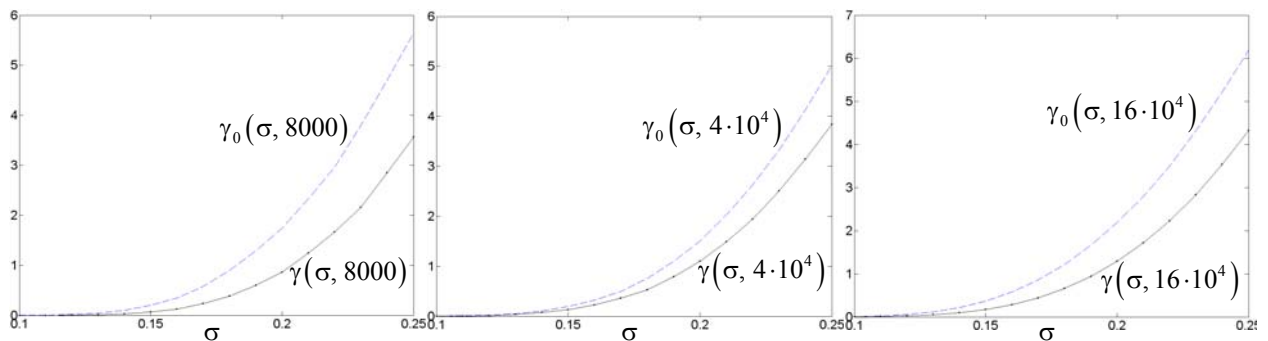


Рис. 10. СПОР нейросети, обученной по множеству (9) при  $C = 3$  и  $F = 8$  для (10) — (12)

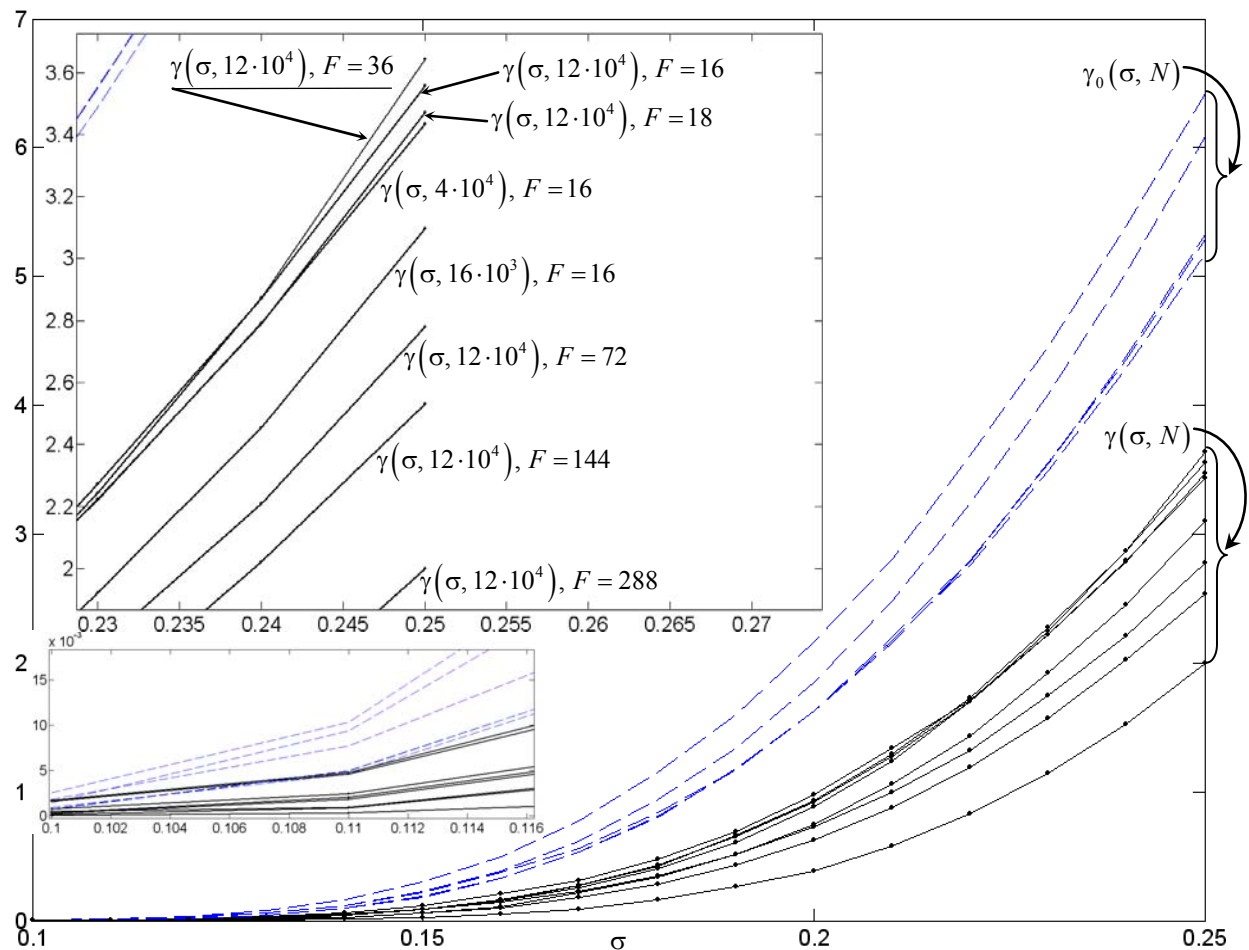


Рис. 11. СПОР нейросети, обученной по множеству (9) при  $C = 2$  и различных  $F$  для (10) — (12)

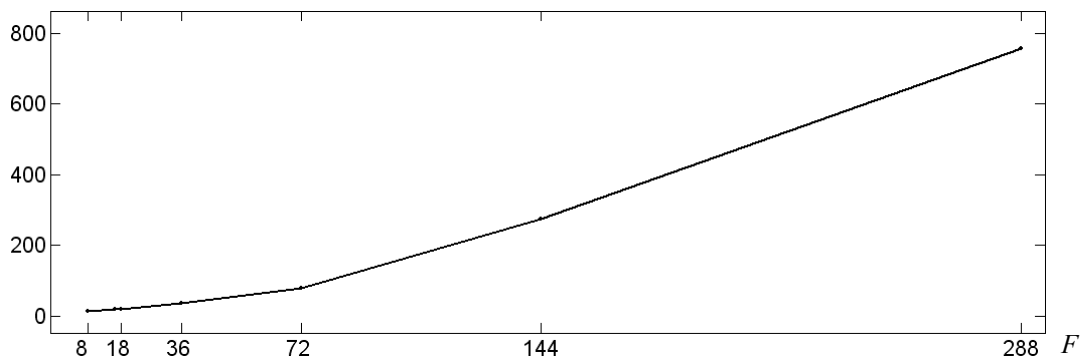


Рис. 12. Ріст середнього часу навчання нейронної мережі для розпізнавання 26 монохромних  $7 \times 5$ -зображень з збільшенням числа  $F$

Як видно, продуктивність нейронної мережі залежить від гладкості навчання, ідентифікованій по частоті розбиття  $F$  передбачуваного відрізка середнього квадратичного відхилення нормального шуму. Ця залежність виражається, перш за все, в практично лінійному зростанні часу навчання і незначительному зменшенні СПОР при збільшенні  $F$ . Такий висновок в більш загальній формулюванні можна розповсюджувати і на інші задачі розпізнавання монохромних зображень з використанням нейронних мереж з прямою зв'язкою і з одним прихованим шаром, де продуктивність нейронної мережі буде підвищуватися при згладженні процесу навчання, який при цьому буде затримуватися.

#### Висновок і перспектива продовження дослідження

Звернемо увагу, що результати тестування нейронної мережі, навченої Matlab-методом `traind`, отримані при 10 нейронах в одному прихованому шарі в середовищі Matlab 7.0.1 з використанням оперативної пам'яті в 512 мегабайт. В більш сучасному середовищі Matlab 7.14 і з використанням більшої пам'яті зростання середнього часу навчання нейронної мережі передбачується не таким інтенсивним, як на рис. 12. Тем не менше, загальна картина протиставлення часу навчання і СПОР повинна зберегтися. Вероятно, вона суттєво не зміниться при збільшенні кількості нейронів, хоча СПОР на певному інтервалі цього кількості знизиться. Однак тут слід уважно контролювати процес навчання, так як при зміні початкових параметрів може збільшитися ймовірність паралічу нейронної мережі або ж для навчання на зашумлених зразках не вистачить епох. В перспективі ці аспекти необхідно вивчати, але перш за все варто зайнятися питанням про кількість повторів  $K$  процедур відклику нейронної мережі на входне множинство (9) з елементами (10) при (11) з множинством ідентифікаторів (12). В дослідженнях, результати яких зображені на рис. 1 — 11, взято  $K=10$ . Що вийде при незначительній зміні цього числа з оптимізованою гладкістю навчання, представляє особливий інтерес, оскільки раціонально вибране число повторів  $K$  додатково знизить СПОР і одночасно прискорить процес навчання.

#### Література

1. Хайкін С. Нейронні мережі: повний курс / Хайкін С. — [2-е изд.]. — М. : Издательский дом «Вильямс», 2006. — 1104 с.
2. Рутковская Д. Нейронные сети, генетические алгоритмы и нечеткие системы / Рутковская Д. — М. : Горячая линия — Телеком, 2004. — 452 с.
3. Каллан Р. Основные концепции нейронных сетей / Каллан Р. — М. : Издательский дом «Вильямс», 2001. — 288 с. : ил.
4. Arulampalam G. A generalized feedforward neural network architecture for classification and regression / G. Arulampalam, A. Bouzerdoum // Neural Networks. — 2003. — Volume 16, Issues 5 — 6. — P. 561 — 568.
5. Hagiwara K. Upper bound of the expected training error of neural network regression for a Gaussian noise sequence / K. Hagiwara, T. Hayasaka, N. Toda, S. Usui, K. Kuno // Neural Networks. — 2001. — Volume 14, Issue 10. — P. 1419 — 1429.
6. Hagan M. T. Neural Networks for Control / M. T. Hagan, H. B. Demuth // Proceedings of the 1999 American Control Conference, San Diego, CA. — 1999. — P. 1642 — 1656.
7. Медведев В. С. Нейронные сети. MATLAB 6 / В. С. Медведев, В. Г. Потемкин. — М. : ДИАЛОГ-МИФИ, 2002. — 496 с.
8. Вапник В. Н. Теория распознавания образов / В. Н. Вапник, А. Я. Червоненкис. — М. : Наука, 1974. — 416 с.

Надійшла 17.1.2013 р.  
Рецензент: д. т. н. Сорокатий Р. В.