

УДК 004.8

**Артамонов Иван Сергеевич**, магистр прикладной математики и информатики, г. Москва

**СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ  
ДЛЯ РАСПОЗНАВАНИЯ ЧЕРНО-БЕЛЫХ ИЗОБРАЖЕНИЙ  
ПРЕДМЕТОВ ОДЕЖДЫ**

**Аннотация.** В данной статье рассматривается вопрос выбора подходов и методов, основанных на машинном и глубоком обучении, применяемых для классификации изображений. Решаемой практической задачей является классификация черно-белых изображений 28x28 пикселей, содержащих изображения одежды. Приводится краткий обзор теоретических оснований классификации, применяемых методов глубокого обучения, а также теоретических аспектов оценки качества работы алгоритмов. Проведены эксперименты с использованием различных методов машинного обучения, исследованы возможности различных архитектур нейронных сетей. Выявлен оптимальный из рассмотренных способ классификации.

**Annotation.** This article discusses the issue of choosing approaches and methods based on machine and deep learning used for image classification. The practical problem being solved is the classification of black and white 28x28 pixel images containing clothing images. A brief overview of the theoretical foundations of classification, the applied deep learning methods, and the theoretical aspects of assessing the quality of the algorithms is given. Experiments were conducted using various machine learning methods, and the capabilities of various neural network architectures were investigated. The optimal classification method was identified.

**Ключевые слова:** классификация изображений, нейронные сети, глубокое обучение, предметы одежды, черно-белые изображения

**Keywords:** image classification, neural networks, deep learning, clothing, black and white images

**Введение**

В последние годы алгоритмы и методы искусственного интеллекта стали применяться повсеместно для анализа и обработки объектов со сложными информационными описаниями. Важными и актуальными остаются задачи обработки текстовых, звуковых, графических данных.

Важными являются технологии, позволяющие анализировать изображения. Они могут применяться отдельными компаниями или маркетплейсами, аналитическими центрами государственных органов или отдельными лицами для поиска, категоризации, систематизации товаров, поиска похожих товаров, мониторинга рынка.

### **Теоретические основания**

В настоящее время большую роль играет такой подход к обработке данных, как многослойные полносвязные и сверточные нейронные сети. Обсудим более подробно принцип организации работы полносвязной нейронной сети.

Полносвязная нейронная сеть (многослойный перцептрон) [1] - это серия нелинейных преобразований, проихводимых послойно. Входные данные в виде числовых признаков объекта поступают во входной слой. Выходные параметры должны воспроизводить неизвестную целевую функцию по обучающей выборке согласно принципу минимизации эмпирического риска.

Обсудим сказанное более подробно.

Общим принципом является использование функции потерь для отдельного объекта  $\mathcal{L}(a, x)$ , используя которую можно получить значение функции потерь для всей выборки  $Q(a, X^\ell)$ . Здесь для примера приведена популярная  $\mathcal{L}(a, x)$  вида квадратичной ошибки. Тогда методом обучения  $\mu(X^\ell)$  будет такой выбор модели  $a$  из семейства алгоритмов  $A$ , который минимизирует потерю на выборке. Функционал  $Q(a, X^\ell)$  называется эмпирическим риском.

$$\mathcal{L}(a, x) = (a(x) - y(x))^2$$

$$Q(a, X^\ell) = \frac{1}{\ell} \sum_{i=1}^{\ell} \mathcal{L}(a, x_i)$$

$$\mu(X^\ell) = \operatorname{argmin}_{a \in A} Q(a, X^\ell)$$

Соответственно, нейронная сеть - это определенное семейство алгоритмов  $A$ . В каждом нейроне происходит суммирование значений, пришедших с прошлого слоя и применение к ним определенной функции активации (см Рис. 1)

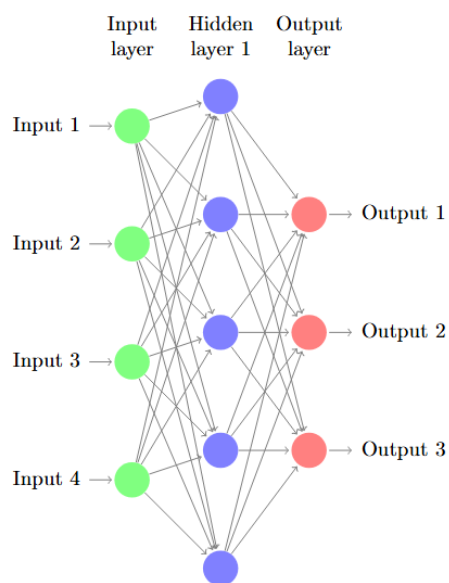


Рисунок 1. Общая схема устройства нейронной сети с одним скрытым слоем

Пусть  $i$ -тый слой выдает значения  $a_i$ . Тогда  $j$ -тый слой будет выдавать значения, выражающиеся следующей формулой:

$$a_j = g_j(\sum w_{ij}a_i + b_j)$$

Параметры  $w_{ij}$  называются весами и на Рисунке 1 могут быть приписаны к прямым линиям, соединяющим нейроны с различных слоёв.

Посредством применения правила дифференцирования сложной функции может быть получено выражение, связывающее производные по весам с различных слоёв нейронной сети. Этот алгоритм называется backpropagation (обратное распространение ошибки)[2].

$$\frac{\partial \mathcal{L}(a, x)}{\partial w_{ij}} = \sum_k \frac{\frac{\partial \mathcal{L}(a, x)}{\partial w_{jk}}}{a_j} w_{jk} g'_j(z_j) a_i$$

Таким образом могут быть вычислены градиенты по всем параметрам модели, и сделан шаг в направлении, противоположном градиенту, то есть, в направлении скорейшего убывания функции потерь.

$$w_{ij} = w_{ij} - \eta \frac{\partial \mathcal{L}(a, x)}{\partial w_{ij}}$$

Не факт, что будет получен набор параметров, соответствующий глобальному минимуму целевой функции потерь, обычно получаемый набор соответственно локальному минимуму.

## Эксперименты

Были проведены эксперименты с датасетом Fashion MNIST, содержащим черно-белые изображения предметов одежды, отшкалированные в квадратах размером 28x28 пикселей (см. Рис.2).

Были использованы нейросети со сверточными[4] и полносвязными слоями а также оптимизатором Adam[3].

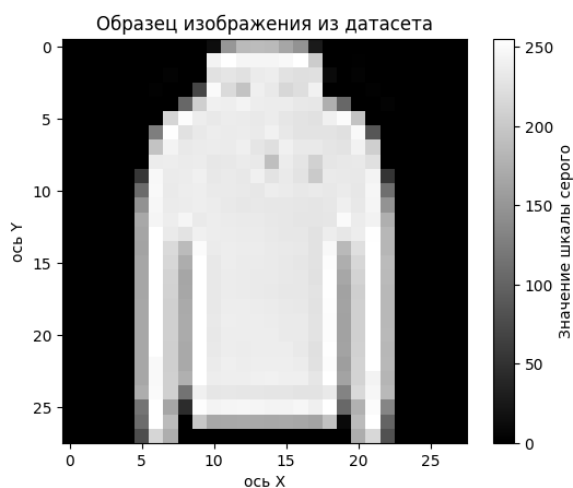
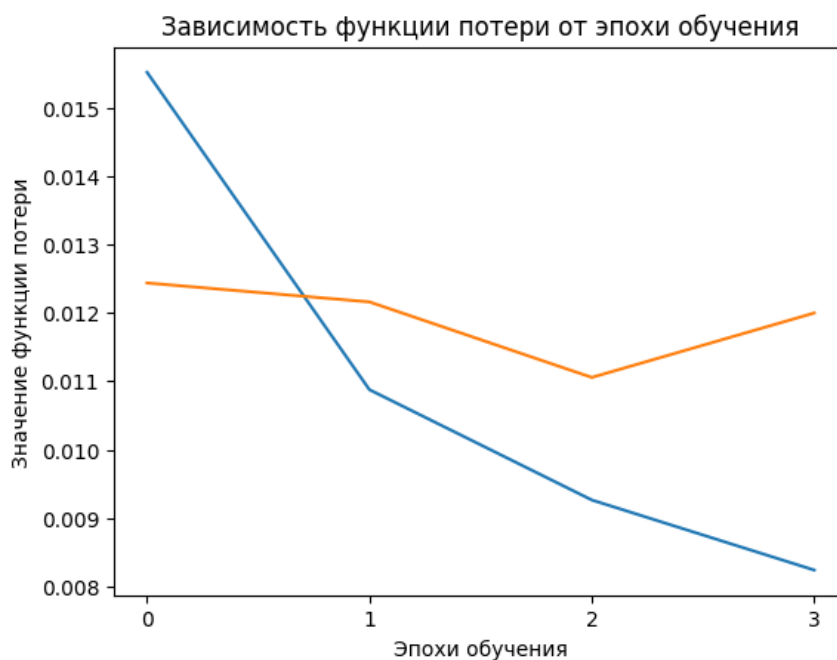


Рис.2 Пример изображения из датасета

Обучающий датасет был разделен на обучающую и валидационную подвыборки для прекращения обучения во избежание переобучения по принципу Early Stopping (прекращение обучения при росте функции потерь на отдельной выборке).



Итоговое значение ассигасы - отношения числа верно классифицированных объектов к общему числу объектов на тестовой выборке составило для разных архитектур, отличающихся в основном числом полносвязных слоёв, следующих после сверточных и слоев max-pooling.

Общая архитектура [1-3 обозначим как CNN]:

1. Сверточный слой с ядром размера 5x5, 6 выходных каналов
2. MaxPooling[5] 2x2 со значением stride=2
3. Сверточный слой с ядром размера 5x5, 16 выходных каналов
4. Многослойная полносвязная сеть

Архитектура	Accuracy
CNN, 1024x120, ReLU, 120x80, ReLU, 80x10	88.94
CNN, 1024x80, ReLU, 80x10	87.19
CNN, 1024x10	85.92
CNN*, 512x10	84.56

CNN\*Слой 3. Имеет не 16, а 8 выходных каналов

Как видно, первая, наиболее сложная архитектура имеет наиболее высокое качество работы, хотя эта разница невелика.

### Заключение

Таким образом, можно прийти к выводу о том, что даже очень простые модели, основанные на сверточных нейронных сетях могут решать данную практическую задачу с качеством, сравнимым с более сложными архитектурами, имеющими значительное число параметров в полносвязных сетях.

Такого рода наблюдения за влиянием числа параметров на качество модели могут быть полезны при многокритериальной оптимизации: затрачиваемая память или время на обучение могут быть в некоторых задачах более важными, чем незначительная потеря в качестве.

Дальнейшие исследования могут включать в себя исследование причин высокого качества работы модели на данном датасете и исследование качества работы модели на других датасетах схожей тематики, установление закономерностей между качеством работы на данном датасете и популярных датасетах, например, CIFAR10.

## **Литература**

1. McCulloch, Warren S., and Walter Pitts. "A logical calculus of the ideas immanent in nervous activity." *The bulletin of mathematical biophysics* 5.4 (1943): 115-133. Z. Zhang, "Improved Adam Optimizer for Deep Neural Networks," 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS), Banff, AB, Canada, 2018, pp. 1-2, doi: 10.1109/IWQoS.2018.8624183.
2. Werbos, Paul. "Beyond regression: New tools for prediction and analysis in the behavioral sciences." PhD thesis, Committee on Applied Mathematics, Harvard University, Cambridge, MA (1974).
3. Z. Zhang, "Improved Adam Optimizer for Deep Neural Networks," 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS), Banff, AB, Canada, 2018, pp. 1-2, doi: 10.1109/IWQoS.2018.8624183.
4. Fukushima, Kunihiko. "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position." *Biological cybernetics* 36.4 (1980): 193-202.
5. LeCun, Yann, et al. "Backpropagation applied to handwritten zip code recognition." *Neural computation* 1.4 (1989): 541-551.

## **Literature**

1. McCulloch, Warren S., and Walter Pitts. "A logical calculus of the ideas immanent in nervous activity." *The bulletin of mathematical biophysics* 5.4 (1943): 115-133. Z. Zhang, "Improved Adam Optimizer for Deep Neural Networks," 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS), Banff, AB, Canada, 2018, pp. 1-2, doi: 10.1109/IWQoS.2018.8624183.
2. Werbos, Paul. "Beyond regression: New tools for prediction and analysis in the behavioral sciences." PhD thesis, Committee on Applied Mathematics, Harvard University, Cambridge, MA (1974).

3. Z. Zhang, "Improved Adam Optimizer for Deep Neural Networks," 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS), Banff, AB, Canada, 2018, pp. 1-2, doi: 10.1109/IWQoS.2018.8624183.
4. Fukushima, Kunihiko. "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position." *Biological cybernetics* 36.4 (1980): 193-202.
5. LeCun, Yann, et al. "Backpropagation applied to handwritten zip code recognition." *Neural computation* 1.4 (1989): 541-551.