

КЛАССИФИКАЦИЯ СЖАТЫХ ИЗОБРАЖЕНИЙ МЕТОДАМИ ГЛУБИННОГО ОБУЧЕНИЯ

П. В. Гончаров

Учреждение образования «Гомельский государственный технический университет имени П. О. Сухого», Республика Беларусь

Классификация изображений – одна из основных задач машинного обучения. Однако использование всех пикселей изображения в качестве входного обучающего вектора делает процесс обучения очень долгим и, вместе с тем, увеличивает количество межнейронных связей, что неизбежно приводит к плохой сходимости функции «стоимости».

В настоящей работе рассматривается метод извлечения признаков изображений, по которым будет производиться дальнейшее распознавание. Как показал М. Крамер [1], нелинейный метод главных компонент (НМГК) работает лучше, чем эквивалентный метод главных компонент (МГК).

Трансформированные в новое пространство признаков изображения нужно классифицировать. Чтобы достичь максимального качества распознавания, следует увеличить количество слоев, что сильно осложняет процесс обучения и снижает вероятность сходимости функции «стоимости». Для решения этой проблемы использована сеть глубокого доверия (СГБ) [2].

Использование СГБ позволяет избежать проблемы взрывного роста или затухания градиентов [3], но процесс обучения СГБ – трудная и тяжелая в вычислительном плане задача. Для ускорения обучения и устранения проблем сходимости в работе применена глубинная нейронная сеть (ГНС) с нормализованной инициализацией весов:

$$W \sim U \left[-\frac{\sqrt{6}}{\sqrt{n_j + n_{j+1}}}, \frac{\sqrt{6}}{\sqrt{n_j + n_{j+1}}} \right], \quad (1)$$

где $U[-a, a]$ – равномерное распределение на интервале $(-a, a)$; n – количество нейронов предыдущего слоя.

Подбор параметров для каждой искусственной нейронной сети осуществлялся с помощью метода *GridSearchCV* (см. *scikit-learn.org*). В таблице представлен сравнительный анализ результатов классификации сжатых до 64 главных компонент изображений (2,5 % от исходного размера изображений для базы *FERET*), из которого видно, что СГБ дает лучший результат. В качестве выборок для обучения и тестов использовались две известных базы данных изображений: «*MNIST handwritten digit database*» и «*FERET face database*».

Сравнительный анализ результатов классификации

Базы	Персептрон	ГНС	СГБ
MNIST	0,7977	0,9285	0,9839
FERET	0,8750	1,0000	1,0000

Литература

1. Kramer, M. A. Nonlinear principal component analysis using autoassociative neural networks // *AIChe journal*. – 1991. – Vol. 37. – № 2. – P. 233–243.
2. Deep belief networks using discriminative features for phone recognition / A. Mohamed [et al.] // 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP). – IEEE, 2011. – P. 5060–5063.
3. Glorot, X. Understanding the difficulty of training deep feedforward neural networks / X. Glorot, Y. Bengio // *Aistats*. – 2010. – Vol. 9. – P. 249–256.