

ИСПОЛЬЗОВАНИЕ НЕЙРОННЫХ СЕТЕЙ В ЗАДАЧЕ РАСПОЗНАВАНИЯ РУКОПИСНЫХ ЦИФР

Шляго Н.А., Сачек И.В., Чичко А.Н.

БГУ, ОАО «Пеленг», Минск, Республика Беларусь, nikitash01@gmail.com

Реферат

Исследовано влияние объема обучающей выборки для различных вариантов нейронных сетей, которые используются для задач классификации рукописных цифр. Показано, что для решения задачи распознавания рукописных цифр наилучшей является нейронная сеть с двумя свёрточными слоями.

Известно, что задача распознавания рукописного текста имеет важное как теоретическое, так и практическое значение [1]. Данная задача решалась различными исследователями: *deep learning* [2], *Q-learning* [2, 3], *SVM* [2, 4], *Random Forest* [5], бустинг [6], что позволяет достигать точности распознавания 80-99%. Так, с помощью *deep learning* достигается точность 99% на основе глубокой свёрточной нейронной сети, а с помощью *SVM (support vector machine)* достигается значение 98%. Также метод *SVM* можно комбинировать со свёрточными нейронными сетями, что позволяет достичь точности в 99.04% [7]. Для решения этой задачи использовались деревья выбора (decision trees) как отдельно, так и в связи со свёрточными сетями [2], а также методы обучения без учителя, и в частности, метод *k* ближайших соседей (*k nearest neighborhoods, kNN*) [3, 8]. Следует отметить, что некоторые вопросы данной задачи в части репрезентативности выборки для задачи распознавания рукописного текста до сих пор не освещены в литературе. В частности, при создании нейронных сетей для малых обучающих выборок не изучено влияние размеров самой выборки [9]. В частности, авторы [2, 3, 4] использовали полные наборы данных, оставляя вопрос о путях уменьшения данного объема выборки неосвещённым.

Целью настоящей работы является решение задачи распознавания рукописных цифр с помощью различных нейронных сетей и определение влияния объема обучающей выборки на качество распознавания символов от «0» до «9».

В качестве исходных данных для нейронных сетей использовали, набор данных MNIST (<http://yann.lecun.com/exdb/mnist/>), содержащий 70000 изображений рукописных цифр размером 28 × 28 пикселей каждая. Используя набор данных MNIST, были сформированы 10 обучающих выборок размерами 100, 1100, 2100, 3100, 4100, 5100, 6100, 7100, 8100, 9100, 10100, 11100, 12100, 13100, 14100, 15100, 16100, 17100, 18100, 19100, 20100 объектов. Причем, в обучающую выборку входило 10 классов, каждый из которых относился к одной цифре от 0 до 9. В качестве тестовой выборки было взято 10000 изображений для каждой выборки. Распределение классов в выборках представлено в таблице 1.

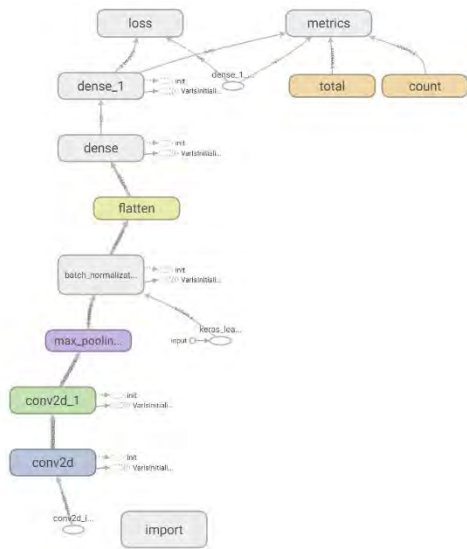
Для решения этой задачи были использованы шесть вариантов нейронных сетей: а) свёрточная нейронная сеть с двумя свёрточными слоями (обозначение варианта НС-1.1 с *BatchNormalization* [10] и варианта НС-1.2 без *BatchNormalization*); б) свёрточная нейронная сеть с одним свёрточным слоем (НС-2.1, НС-2.2); в) полносвязная нейронная сеть с тремя полносвязными слоями (НС-3.1, НС-3.2).

Таблица 1 – Распределение объектов различных типов в выборках

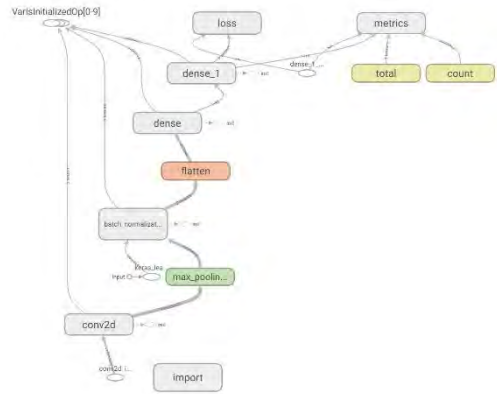
Объём выборки, шт	Количество объектов в обучающей выборке различного типа, шт									
	типа «0»	тип «1»	типа «2»	типа «3»	типа «4»	типа «5»	типа «6»	типа «7»	типа «8»	типа «9»
Обучающая выборка										
100	13	14	6	11	11	5	11	10	8	11

200	21	26	20	21	21	13	19	21	15	23
300	34	39	28	34	32	23	29	29	21	31
400	39	52	42	41	44	30	36	42	30	44
500	50	66	52	50	52	39	45	52	39	55
600	58	79	64	59	59	51	54	62	49	65
700	69	90	76	67	71	62	63	75	55	72
800	79	98	81	73	81	72	73	86	73	84
900	87	106	93	86	95	81	82	98	80	92
1000	97	116	99	93	105	92	94	117	87	100
1100	109	132	109	100	116	99	99	130	98	108
2100	199	230	209	202	226	188	210	235	186	215
3100	294	346	307	304	339	279	316	337	277	301
4100	388	456	406	400	437	360	415	443	383	412
5100	491	574	494	504	546	442	513	560	473	503
6100	601	682	593	621	631	523	616	661	559	613
7100	703	797	684	725	725	618	717	764	659	708
8100	808	917	797	826	807	708	810	862	760	805
9100	905	1033	896	932	907	784	922	971	856	894
10100	1011	1138	1004	1043	984	873	1022	1081	954	990
11100	1114	1251	1100	1149	1096	962	1123	1179	1036	1090
12100	1212	1363	1181	1239	1198	1058	1215	1289	1142	1203
13100	1314	1471	1279	1350	1293	1146	1308	1393	1241	1305
14100	1414	1585	1378	1464	1377	1238	1399	1495	1345	1405
15100	1504	1703	1474	1564	1477	1328	1501	1601	1438	1510
16100	1612	1818	1569	1673	1560	1426	1600	1700	1533	1609
17100	1701	1943	1664	1773	1649	1516	1697	1804	1641	1712
18100	1791	2062	1750	1895	1760	1593	1784	1910	1738	1817
19100	1893	2183	1848	1988	1862	1683	1878	2008	1838	1919
20100	2001	2295	1940	2085	1956	1786	1977	2104	1933	2023
Тестовая выборка										
10000	980	1135	1032	1010	982	892	958	1028	974	1009

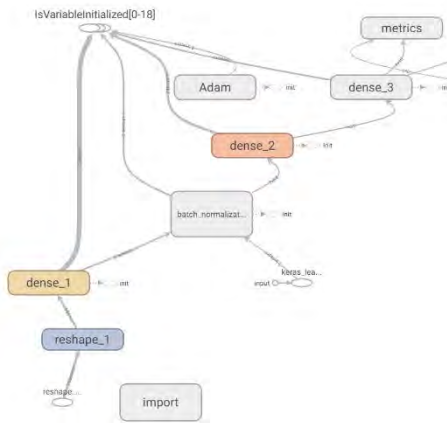
Для построения вышеперечисленных нейронных сетей была разработана программа на языке Python, использующая библиотеки Keras, TensorFlow. На рисунке 1 представлены графы нейронных сетей, которые использовались для решения задачи.



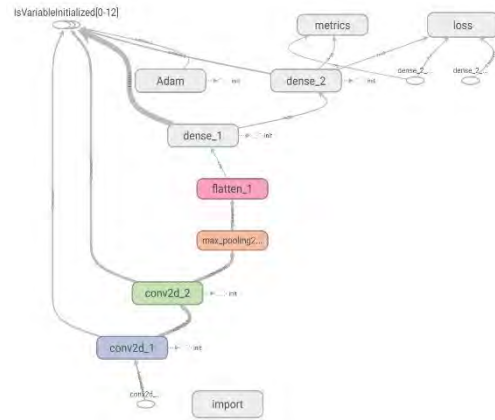
а)



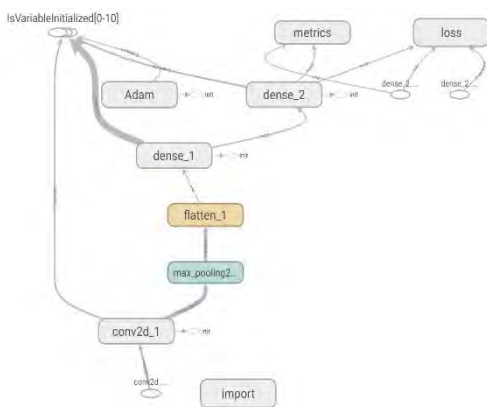
б)



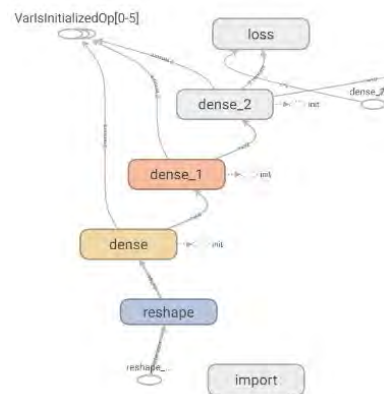
в)



г)



д)



е)

Рисунок 1 – Графы нейронных сетей, использованные для решения задачи распознавания рукописных цифр: а) НС-1.1; б) НС-2.1; в) НС-3.1; г) НС-1.2; д) НС-2.2; е) НС-3.2

Для исследования было проведено четыре эксперимента: данные не нормализовались, нормализация в сети отсутствовала (вариант 1, сети НС-1.2, НС-2.2, НС-3.2); данные не нормализовались, нормализация в сети присутствовала (вариант 2, сети НС-1.1, НС-2.1, НС-3.1); данные нормализовались, нормализация в сети отсутствовала (вариант 3, сети НС-1.2, НС-2.2, НС-3.2), данные нормализовались, нормализация в сети присутствовала (вариант 4, сети НС-1.1, НС-2.1, НС-3.1). Во всех вариантах задачи в качестве критерия качества распознавания объектов использовался показатель, который рассчитывался по формуле: $k = \frac{T}{F+T} \times 100\%$, где T – количество правильно распознанных объектов в тестовой выборке, F – количество неправильно распознанных объектов. На рисунках 2 и 3 и в таблице 2 представлены данные результатов распознавания, полученные для малых и больших объемов выборок для вариантов задачи 1-4.

Как видно из рисунка 2 (вариант задачи 1) показатели качества распознавания рукописных цифр (вариант НС 1.2) для выборки, состоящей из 3000 объектов превышают 85 %. Для вариантов нейронных сетей НС-2.2; 3-НС-3.2 наблюдается колебание показателя качества, что свидетельствует о неустойчивости распознавания рукописных цифр. На малых объемах выборки показатель качества распознавания является неустойчивым для сетей НС-1.2. 2-НС-2.2. 3-НС-3.2 (рисунок 3).

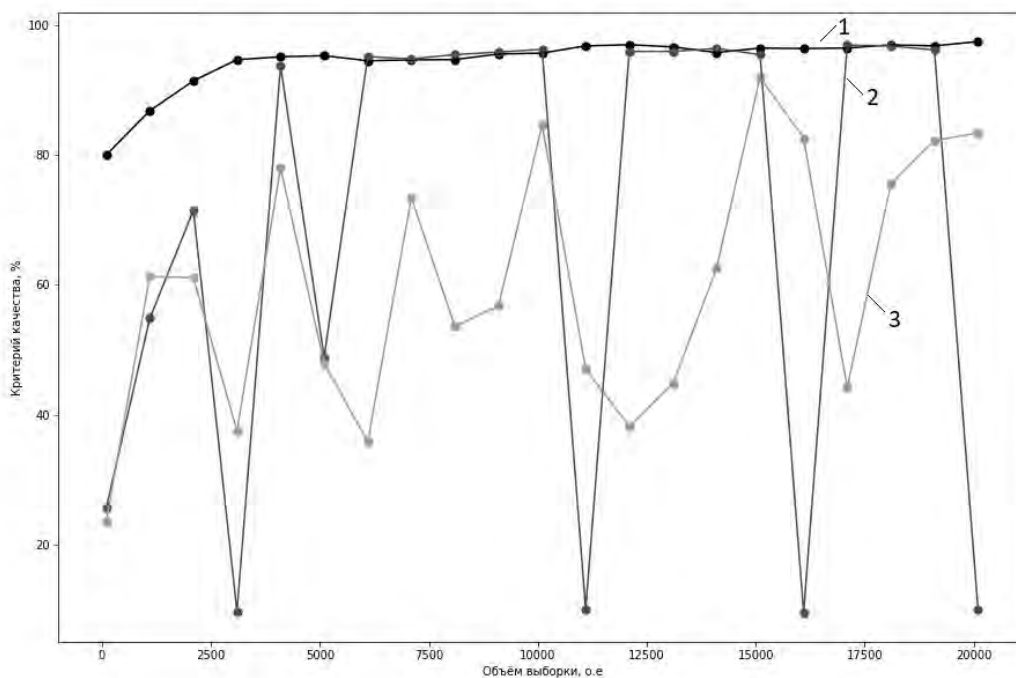


Рисунок 2 – Влияние объёма выборки на критерий качества распознавания рукописных цифр для различных нейронных сетей при ненормализованных данных (вариант задачи 1): 1-НС-1.2; 2-НС-2.2; 3-НС-3.2

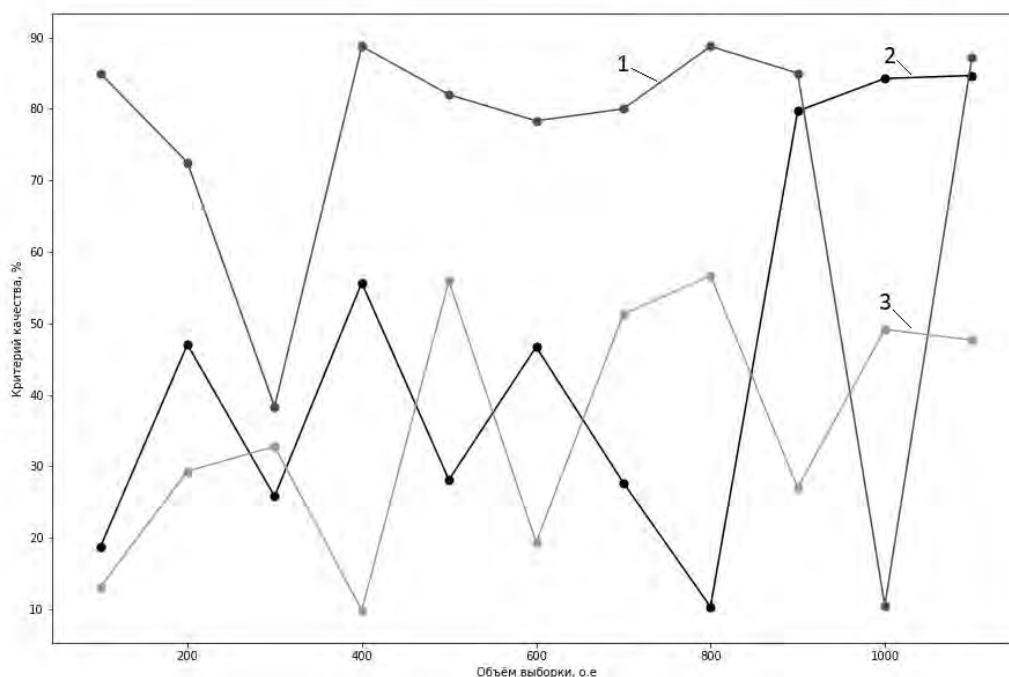


Рисунок 3 – Влияние малых объёмов выборки (100-1000 о.е.) на критерий качества распознавания рукописных цифр для различных нейронных сетей при ненормализованных данных (вариант задачи 1): 1-НС-1.2; 2-НС-2.2; 3-НС-3.2

Таблица 2 – Влияние объёма выборки на критерий качества распознавания рукописных цифр для вариантов задачи 2, 3 и 4

Размер выборки	Вариант 2			Вариант 3			Вариант 4		
	НС-1.1	НС-2.1	НС-3.1	НС-1.2	НС-2.2	НС-3.2	НС-1.1	НС-2.1	НС-3.1
100	68,21	59,78	52,89	67,93	68,8	53,26	66,95	71,05	55,01
200	76,98	90,91	63,02	79,38	77,26	68,39	79,28	77,33	65,66
300	81,25	83,33	71,24	81,46	80,44	71,65	83,65	80,32	72,35
400	86,25	87,10	77,83	83,75	81,51	77,15	85,05	85,06	78,17
500	87,15	89,02	79,75	86,85	84,69	81,86	87,02	86,75	80,43
600	89,92	87,25	82,45	86,33	87,37	83,82	90,45	89,09	83,4
700	89,97	82,79	84,75	87,59	88,72	84,96	91,65	89,95	83,5
800	91,49	88,73	84,81	88,95	88,64	86,64	92,08	90,45	84,54
900	92,31	90,12	85,1	89,75	89,42	86,71	92,66	91,02	84,37
1000	91,33	86,81	85,99	91,21	89,7	87,06	93,25	91,9	86,49
1100	92,95	91,72	86,93	92,01	89,87	87,48	93,89	91,71	86,59
2100	95,20	94,12	89,62	95,17	92,86	90,75	95,53	94,03	89,56
3100	96,42	95,13	91,53	95,39	94,24	91,42	96,67	95,64	91,84
4100	97,18	95,94	92,64	96,31	95,47	91,55	96,58	96,26	92,33
5100	97,57	96,66	93,07	96,44	95,48	92,88	97,51	96,53	92,79
6100	97,52	96,71	93,29	96,76	96,35	93,44	97,65	97,07	93,38
7100	98,01	96,99	94,0	97,3	96,68	93,8	97,44	97,23	93,73
8100	97,34	97,36	94,23	97,15	96,62	94,05	97,77	97,35	94,23
9100	97,82	97,21	94,22	97,11	97,04	94,21	97,73	96,83	93,93

10100	97.54	97.0	94.67	97.53	97.26	94.51	97.87	97.19	94.39
11100	98.21	96.69	94.53	97.5	97.17	94.41	98.02	96.93	95.15
12100	98.07	96.62	94.87	97.43	97.32	94.82	97.41	97.26	94.92
13100	98.08	97.14	94.83	97.58	97.64	95.19	97.49	97.33	95.07
14100	98.13	97.48	95.47	97.95	97.62	95.43	97.6	97.4	95.3
15100	97.64	97.29	95.54	97.94	97.65	95.71	97.83	97.75	95.53
16100	98.43	97.56	95.78	98.14	97.62	95.8	98.27	97.57	95.91
17100	98.02	97.33	95.72	98.16	97.63	95.89	98.12	97.71	95.89
18100	98.26	97.56	95.93	98.28	97.72	95.91	98.04	97.67	95.94
19100	98.09	97.57	96.02	98.21	97.99	95.86	98.41	97.6	96.14
20100	98.23	97.6	96.23	98.36	97.95	95.99	98.24	97.76	96.26

Как видно из таблицы 2, показатель качества колеблется в интервале 84-99% для нейронных сетей в вариантах задачи 2, 3, 4 при объемах выборок свыше 900. Следует отметить что при объеме выборки находящейся в интервале 3500 – 4500 объектов показатель качества вступает в фазу насыщения (близкое к постоянному значению критерия качества) вне зависимости от нормализации входных данных. Также видно, что нормализацию данных можно заменить на внутреннюю нормализацию в первых слоях сети и что комбинация «нормализованные данные и *BatchNormalization*» не даёт увеличения критерия качества.

Таким образом, в рассматриваемых сетях можно делать что-то одно: либо нормализовать данные, либо добавлять слой *BatchNormalization*. По-видимому, разные нейронные сети в зависимости от их морфологии имеют различные уровни минимального размера выборок, для которых наблюдается наилучшее распознавание. Причем свёрточные сети по вычислительным характеристикам и точности более предпочтительны, так как результаты получаются более стабильными, а также результаты нейронной сети с двумя свёрточными слоями не зависят от нормализации данных. Следовательно, для данного класса задач из 3-х выбранных нейронных сетей нейронные сети с двумя свёрточными слоями являются наилучшими для распознавания рукописных цифр.

Заключение

Показано, что для решения задач распознавания рукописных цифр наилучшей является нейронная сеть с двумя свёрточными слоями. Для различных обучающих выборок вычислены показатели качества распознавания. Показано, что увеличение для сверточной сети объема выборки свыше 4000 объектов не приводит к улучшению качества распознавания. Показано, что результаты нейронной сети с двумя свёрточными слоями не зависят как от нормализации данных, так и от наличия внутренней нормализации при объемах выборки свыше 1000, в то же время для двух других исследованных сетей наблюдается зависимость показателя качества от нормализации данных.

Литература

1. И.Л. Кафтанников, А.В. Парасич Проблемы формирования обучающей выборки в задачах машинного обучения // Южно-Уральский государственный университет, г. Челябинск.
2. Shashank Mishra, 2D. Malathi, 3K. Senthilkumar Digit recognition using deep learning // International Journal of Pure and Applied Mathematics, 2018, V.118. - №22. – pp. 295-302.
3. Junfei Qiao, Gongming Wang, Wenjing Li, MinChen An adaptive deep Q-learning strategy for handwritten digit recognition // Neural Networks, 2018, V.107, pp. 61-71.
4. Anuj Dutt, Aashi Dutt Handwritten Digit Recognition Using Deep Learning // International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 2017, V.6, pp. 990-997.

5. Simon Bernard, Laurent Heutte, Sébastien Adam Using Random Forests for Handwritten Digit Recognition // <https://hal.archives-ouvertes.fr/hal-00436372>, 2009.
6. Eva Volna, Vaclav Kocian and Martin Kotyrba Boosting of Neural Networks over MNIST Data // Proceedings of the International Conference on Neural Computation Theory and Applications (NCTA-2014), pp. 256-263.
7. Abien Fred M. Agarap An Architecture Combining Convolutional Neural Network (CNN) and Support Vector Machine (SVM) for Image Classification // arXiv:1712.03541v2 [cs.CV] 7 Feb 2019.
8. Shengfeng Chen, Rabia Almamlook, Yuwen Gu, Dr. Lee wells Offline Handwritten Digits Recognition Using Machine learning // Proceedings of the International Conference on Industrial Engineering and Operations Management Washington DC, USA, September 27-29, 2018
9. А.В. Савчик, Е.А. Швец, Д.П. Николаев Разработка архитектуры и методов обучения обобщающих нейросетевых классификаторов изображений при малом размере обучающей выборки // Информационные процессы, 2017, Т.17, №4, стр. 309-317.
10. Sergey Ioffe, Christian Szegedy Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift // arXiv:1502.03167v3 [cs.LG] 2 Mar 2015.