Информационные процессы, Том 17, № 4, 2017, стр. 309–317.

МАТЕМАТИЧЕСКИЕ МОДЕЛИ, ВЫЧИСЛИТЕЛЬНЫЕ МЕТОДЫ

Разработка архитектуры и методов обучения обобщающих нейросетевых классификаторов изображений при малом размере обучающей выборки [1]

А.В. Савчик, Е.А. Швец, Д.П. Николаев

Институт проблем передачи информации, Российская академия наук, Москва, Россия
email: savsmail@gmail.com, shvets@iitp.ru, dimonstr@iitp.ru

Поступила в редколлегию 21.11.2017

Аннотация—В работе рассматривается метод обучения нейросетевого классификатора при малом размере обучающей выборки с привлечением дополнительной выборки данных большего размера. Метод заключается в обучении обобщающей искусственной нейронной сети древовидной структуры, первые (префиксные) слои которой являются общими для нескольких (в исследуемом случае – двух) задач классификации, каждый лист является выходным слоем классификатора только для одной из задач, а полный классификатор для каждой из задач состоит из последовательности слоев, ведущих от основания дерева к этому листу. Предлагаемый метод апробирован на примере задачи распознавания символов номерных знаков при размере обучающей выборки, равной одному элементу на класс. Выполнено сравнение полученных результатов с результатами систем, обученных на малой обучающей выборке другими методами.

КЛЮЧЕВЫЕ СЛОВА: машинное обучение, искусственные нейронные сети, обобщающее обучение, классификация изображений

1. ВВЕДЕНИЕ

Искусственные нейронные сети (в частности, сети сверточной архитектуры) получили широкое распространение в области классификации изображений [1, 2, 3]. Одной из основных проблем при обучении таких сетей является создание обучающей выборки (“датасета”) достаточного объема и репрезентативности (то есть покрывающей множество примеров, которые будут подаваться на вход нейросетевому классификатору в процессе эксплуатации), поскольку такая выборка может состоять из десятков тысяч элементов [4, 5], и при ее создании необходимо с помощью эксперта определять (или хотя бы проверять) метки для каждого элемента. Иногда обучающая выборка может быть получена только в процессе работы уже обученной системы, а также бывают случаи, когда невозможно проактивно осуществлять сбор данных (например, данные о сейсмической активности перед землетрясением).

Альтернативой созданию обучающей выборки является использование одного из общедоступных датасетов. Однако такие датасеты существуют не для всех задач; и даже для задач, для которых датасеты доступны в открытом доступе (например, для распознавания символов алфавита [5]), они обычно не могут быть единственным источником данных, поскольку не содержат изображений, специфичных именно для решаемой, целевой задачи (например, при распознавании символов общедоступный датасет скорее всего не будет содержать символов необходимых шрифтов). В данной работе мы предлагаем использовать общедоступные

[1] Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта №16-07-01167.

датасеты в качестве дополнительного источника данных для обучения нейросетевых классификаторов изображений при малом размере целевой выборки.

Предлагаемый подход заключается в обучении искусственной нейронной сети (ИНС) древовидной структуры (рис. 1). Обучение производится одновременно с использованием двух обучающих выборок: целевой (малого размера) и дополнительной (большого размера). Такая сеть учится решать две задачи классификации, и ее можно представить как две отдельные ИНС, которые имеют общие слои и обучаются одновременно.

В работе мы применяем предложенный метод для случая, когда обучающая выборка для целевой задачи содержит 1 элемент на класс. Проблема обучения нейросетевого классификатора при размере обучающей выборки в 1 элемент на класс широко исследуется и получила в англоязычной литературе название “one-shot learning” [6, 7]. Одним из классических подходов к этой проблеме является метод обучения переносом (“transfer learning”) [3, 8]. Метод обучения переносом, как и предлагаемый нами, предполагает наличие дополнительного датасета. Пример обучения ИНС методом переноса заключается в том, что сначала обучается ИНС, классифицирущая изображения дополнительного датасета; затем обучается вторая ИНС с использованием целевого датасета, но при этом несколько первых (префиксных) слоев второй ИНС берутся без изменений из первой (обученной на дополнительном датасете) ИНС и не изменяются в процессе дообучения на целевых данных.

Другим подходом, родственным предлагаемому нами, является метод многозадачного обучения (“multi-task learning”) [9]. В этом случае ИНС обучается решать несколько тесно связанных задач – например, определять одновременно положение глаз, носа или рта человека и выражение его лица [10]). При этом задачи как правило являются одинаково важными и решаются на одних и те же данных (для приведенного выше примера позиция носа и выражение лица определяется на одной и той же фотографии). Аналогично данному методу, обучаемая нами ИНС решает несколько задач, однако мы оперируем с несколькими датасетами, не накладываем жестких ограничений на связанность задач и имеем только одну целевую задачу.

Заслуживающим упоминания способом предобработки данных малого датасета является его искусственное увеличение за счет искажения имеющихся примеров – т.н. “аугментация данных” [1, 11]. Обратим внимание, что для эффективного использования аугментации необходимо априорное знание о возможных искажениях изображений, которые могут встретиться системе в процессе эксплуатации.

2. ОБОЩАЮЩИЙ НЕЙРОСЕТЕВОЙ КЛАССИФИКАТОР ИЗОБРАЖЕНИЙ

2.1. Архитектура обобщающей ИНС

Предлагаемый нами метод заключается в обучении древовидного классификатора (рис. 1), каждый лист которого соответствует выходному слою одной из решаемых задач, а полный классификатор для решения задачи состоит из последовательности слоев, ведущих от основания дерева к соответствубщему листу. Несколько первых слоев являются общими для классификаторов разных задач, что позволяет обнаружить и перенести информацию об особенностях исходных выборок с дополнительного на целевой классификатор и предотвратить процесс переобучения на малом числе целевых данных. При этом обучение классификатора производится одновременно с использованием обоих датасетов.

Такой метод похож на метод обучения ИНС путем переноса знаний (“transfer learning”), вариантом которого является обучение нескольких “префиксных” слоев нейросетевого классификатора с использованием дополнительного датасета, и после этого обучения оставшихся слоев на малом, целевом датасете. В отличие от нашего метода, такой подход вообще не ис-

пользует целевую выборку при обучении первых слоев ИНС, следовательно, префиксные слои обученного классификатора являются неоптимальными для целевой задачи.

Таблица 1. Слои нейросетевого классификатора изображений

Тип слоя	Размер ядра	Число выходных каналов/нейронов	Общий слой
Вход	–	11	
Сверточный	5*5	4	Да
Max. pooling	2*2	4	Да
Сверточный	3*3	8	Да
Max. pooling	2*2	8	Да
Сверточный	3*3	16	Да
Сверточный	1*1	16	Нет
Полносвязный	–	24 (буквы) или 10(цифры)	Нет
Softmax	–	24 (буквы) или 10(цифры)	Нет

На вход сети подавались одноканальные изображения размером 16 на 16 пикселей. Архитектура ИНС, использованная в экспериментах, приведена в таблице 1. Число раздельных слоев классификаторов в экспериментах варьировалось, но во всех экспериментах обе ветки классификатора имели одинаковую структуру (кроме последнего слоя, число выходных нейронов которого равно числу классов соответствующей задачи).

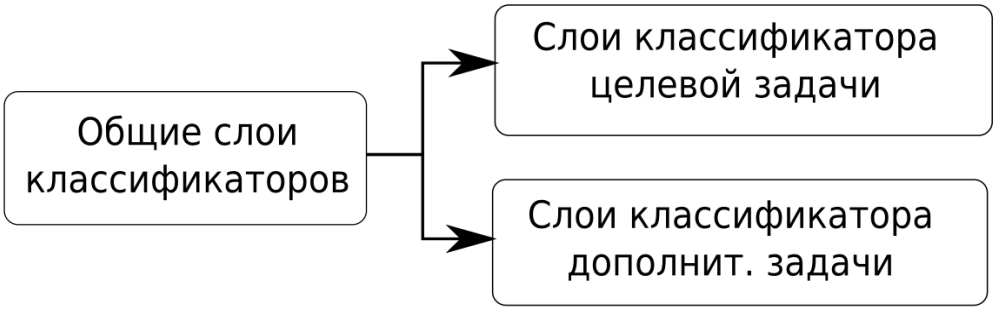

Рис. 1. Архитектура обобщающего нейросетевого классификатора

2.2. Метод обучения обобщающей ИНС

В процессе обучения обобщающей ИНН участвуют оба датасета. При обучении мы использовали метод стохастического градиентного спуска и функцию кросс-энтропии в качестве функции потерь (поэтому скорость обучения стохастического градиентного спуска зависела от числа элементов в минибатче (подвыборке, по которой считается итерация стохастического градиента), подаваемых в систему за одну эпоху обучения, а не от числа элементов в каждом из них). Обозначим N_b число минибатчей дополнительной и N_g число минибатчей целевой выборки выборки. В нашем случае использовалось $N_b = 20$ и $N_g = 1$.

Обучение состояло из последовательных итераций на дополнительных и целевых данных. Чтобы регулировать, насколько важным является каждый из датасетов, для разных датасетов были установлены отличающиеся значения коэффициента скорости обучения: $\lambda_{\text{цел}}$ для целевого и $\lambda_{\text{доп}}$ для дополнительного. Будем называть относительной скоростью обучения параметр $s = \frac{N_g \lambda_{\text{цель}}}{N_b \lambda_{\text{доп}}}$.

При обучении использовался критерий останова по числу эпох ($N = 2000$). Этот критерий выбран экспериментально и соответствует времени обучения, при котором большинство обу-

чаемых ИНС достигли достаточно высокого качества классификации без ярко выраженного переобучения.

3. ДАННЫЕ ДЛЯ ОБУЧЕНИЯ

При постановке экспериментов были использованы три источника данных. Основная часть экспериментов была проведена на датасете, состоящем из цифр и букв автомобильных номеров, собранном в рамках проекта по распознавания регистрационных номеров автомобилей [12]. В качестве целевой выборки выступали буквы номеров, при этом в обучающую выборку случайно выбиралась 1 элемент на класс.

В процессе экспериментов три различные обучающие выборки использовались в качестве дополнительной:

- Цифры регистрационных номеров автомобилей [12].
- Датасет рукописных букв MNIST [13].
- База размеченных символов номеров домов SVHN ('Street View House Number Dataset') [2].

Если не сказано иначе, в экспериментах использовалась первая выборка в качестве дополнительной. Примеры изображений целевой и дополнительных выборок приведены на рисунке 2.

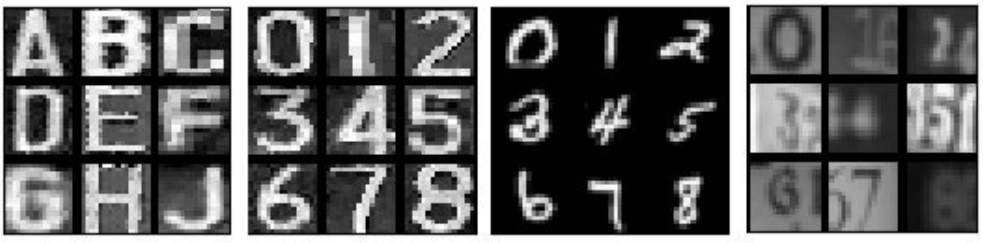

Рис. 2. Примеры изображений использованных обучающих выборок, слева направо: буквы и цифры регистрационных номеров автомобилей, MNIST, SVHN

4. ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ

4.1. Влияние коэффициентов скорости обучения на качество классификатора

Первая серия экспериментов была проведена с целью исследовать, как коэффициенты скоростей обучения и относительная скорость обучения s влияет на качество обученного классификатора на целевом датасете.

Для этого были обучены несколько обобщающих классификаторов с архитектурой, приведенной в таблице 3, с двумя разделенными слоями. Коэффициент скорости обучения составил $\lambda_{\text{доп}} = 0.01$ для дополнительного датасета и принимался значения из набора $\lambda_{\text{цел}} \in \{0.006, 0.02, 0.06, 0.2, 0.6\}$ для целевого датасета. Во всех экспериментах, описанных в этом и следующих параграфах, для каждой из обучаемых ИНС элементы целевой выборки были выбраны случайно 5 разными способами, а качество полученных ИНС усреднены.

Полученные результаты представлены в таблице 4.1. Как следует из представленных результатов, скорость обучения на целевой выборке должна быть достаточна мала по сравнению со скоростью обучения на дополнительной выборке для предотвращения переобучения.

$\lambda_{\text{цел}}$	s	качество без доп. выборки	качество с доп. выборкой
0.006	0.03	0.423 ±0.0146	0.639 ±0.039
0.02	0.1	0.422 ±0.0595	0.648 ±0.051
0.06	0.3	0.453 ±0.0672	0.619 ±0.042
0.2	1.0	0.449 ±0.0654	0.562 ±0.053
0.6	3.0	0.455 ±0.052	0.57 ±0.041

Таблица 2. Результаты обучения для различных скоростей обучения

4.2. Влияние числа раздельных слоев на качество классификатора и сравнение с обучением переносом

В таблице 3 приведена зависимость итогового качества классификатора на целевой выборке от числа раздельных слоев. Классификаторы были обучены двумя способами: обучением обобщающего классификатора и путем обучения переносом (в этом случае сначала обучалась ИНС на дополнительном датасете, после этого коэффициенты первых нескольких слоев фиксировались, а оставшиеся слои доо,учались на целевом датасете).

Из результатов следует, что для работы обучающего классификатора на рассмотренных данных достаточно одного раздельного слоя, а добавление лишних слоев ведет к снижению качества за счет переобучения. В то же время обучение переносом дает низкое качество классификатора, которое растет с уменьшением числа слоем, которые фиксируются перед обучением на целевой обучающей выборке. Это свидетельствует о неоптимальности коэффициентов первых слоев при использовании этого метода.

Таблица 3. Результаты обучения при различном числе общих слоев и методах обучения

число разд. слоев	метод обучения	качество классификатора
без доп. выборки	–	0.422 ±0.059
1	обобщающий	0.697 ±0.059
1	переносом	0.464 ±0.055
2	обобщающий	0.648 ±0.051
2	переносом	0.479 ±0.057
3	обобщающий	0.563 ±0.057
3	переносом	0.504 ±0.044

4.3. Использование аугментации данных для обучения ИНС на малой выборке

Аугментация данных заключается в добавлении в обучающую выборку примеров, преобразованных из имеющихся, и снабженных метками, следующими из априорных знаний о генеральной совокупности (в большинстве случаев аугментация сохраняет метки искажаемых примеров).

Мы исследовали, как использование аугментации и метода обобщающего обучения влияет на точность итогового классификатора. Процесс аугментации в нашей работе носит вероятностный характер: к каждому изображению с некоторыми вероятностями применяются искажения зашумления пикселей ($I'(x,y) = I(x,y) + \Delta I(x,y)$), контрастирования случайной монотонной функцией (генерирующейся как интеграл последовательности случайных величин), поворота, сдвига и скоса (преобразование, сохраняющее y и преобразующее координату x по закону $x' = x + \alpha y$). Вероятности искажений приведены в таблице 4 (в случае обрезанных нормальных распределений интервал, на котором распределение отлично от нуля, равен двум σ), а результаты обучения – в таблице 5. Как следует из таблицы, качество системы, обученной на аугментированных данных, немного превышает качество системы, обученной методом обобщающего обучения, но именно использование обоих методов позволяет добиться максимального качества – при аугментации выборки в 100 раз и использовании обобщающего обучения итоговое качество классификатора составило 0.849 ± 0.044, по сравнению с каче-

ством 0.709 ± 0.034 при использовании только аугментации и качеством 0.423 ± 0.0146, когда не используется ни аугментации, ни обобщающего обучения.

Таблица 4. Параметры аугментации данных

Искажение	вероятность	параметры	распределение
Зашумление	0.8	$\Delta I(x,y) \propto N(0, \sigma = .05)$	Обрезанное нормальное
Контрастирование	0.8	–	–
Поворот	0.8	$\alpha \in [-3, 3]$	Равномерное
Сдвиг	0.8	$\Delta\rho \in [-1, 1]$	Равномерное
Скос	0.8	$\alpha \propto N(0, 0.2)$	Обрезанное нормальное

Таблица 5. Результаты обучения с использованием аугментации и обобщающего обучения

	Качество обученных классификаторов	
Коэффициент аугментации	Аугментация	Аугментация + обобщающий классификатор
2	0.553±0.061	0.667±0.054
10	0.661 ±0.049	0.76±0.052
100	0.709 ±0.034	0.849±0.044

4.4. Исследование качества обученного классификатора для различных дополнительных датасетов

Интуитивно кажется, что качество обученного обобщенного классификатора будет тем выше на целевой выборке, чем больше дополнительный датасет “похож” на целевой. Для рассматриваемой в работе задачи (распознавание букв номеров) достаточно легко найти общедоступный похожий датасет, но для некоторых задач классификации может оказаться, что все доступные дополнительные датасеты для обучения слабо скореллированы с целевым и собраны из другой предметной области. Поэтому интересным является вопрос, как меняется качество целевого классификатора при использовании различных дополнительных датасетов. Мы исследовали три разных дополнительных датасета, описанных в разделе 3. Результаты приведены в таблице 6. Интересно, что хотя датасет цифр наиболее близок к целевому датасету букв (символы вырезаны из общего датасета регистрационных номеров автомобилей), при использовании датасета рукописных символов MNIST в качестве дополнительного качество классификатора оказалось практически идентичным – вероятно, за счет уменьшения эффекта переобучения. При этом дополнительная выборка символов с табличек с номерами домов оказалась менее эффективной (частично из-за использования неудачного критерия останова для этого случая).

Таблица 6. Результаты обучения при использовании различных дополнительных датасетов

Дополнительный датасет	качество классификатора
Отсутствует	0.422 ±0.0595
Цифры номеров	0.648 ±0.0511
MNIST	0.643 ±0.0365
SVHN	0.511 ±0.0561

4.5. Исследование сходимости обучения классификатора на целевом и дополнительном датасетах

При обучении нейронных сетей необходимо осуществлять остановку обучения до того, как сеть вошла в переобучение [14]. Эффективные критерии останова обучения обычно определяют момент останова обучения путем анализа качества классификатора на валидационной

выборке. При обучении же на малой выборке (в особенности при размере выборки 1 элемент на класс) элементов для создания валидационной выборки недостаточно, потому встает вопрос, как останавливать обучение обобщающей сети.

В данной работе мы использовали обучение в течение фиксированного числа эпох. Для дополнительных датасетов цифр и рукописных символов при таком подходе не возникло переобучения на целевой выборке, при использовании же символов с номеров домов ИНС вошла в режим переобучения примерно через 1000 эпох. Графики, иллюстрирующие процесс обучения, приведены на рисунках 3, 4, 5.

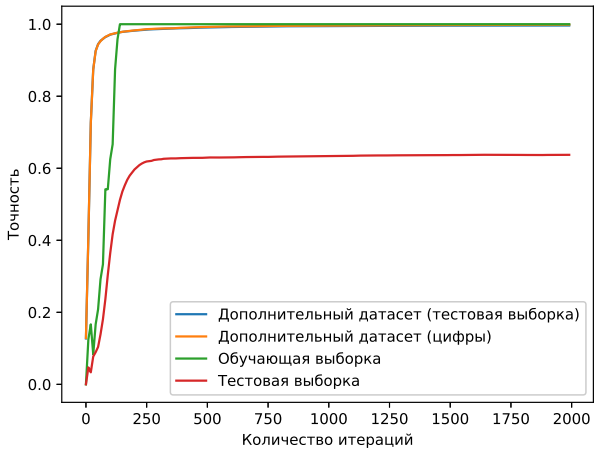

Рис. 3. Графики обучения (доп. датасет - цифры)

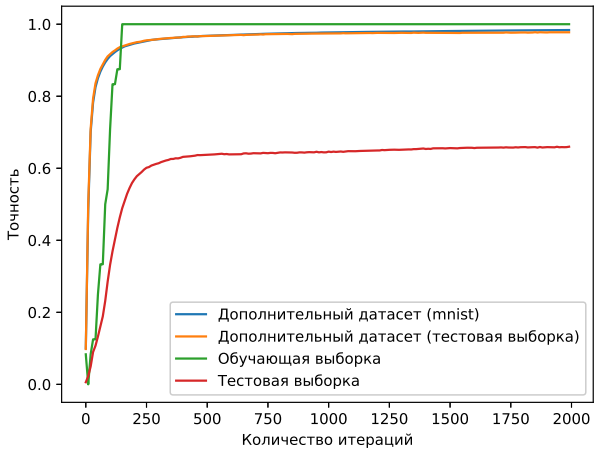

Рис. 4. Графики обучения (доп. датасет - MNIST)

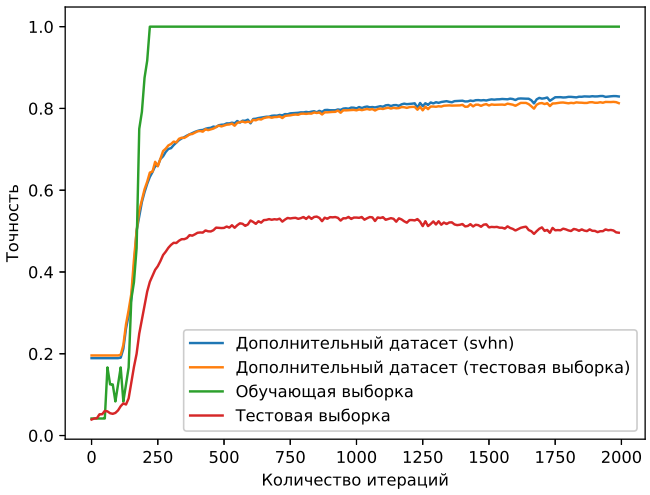

Рис. 5. Графики обучения (доп. датасет - SVNH)

5. ЗАКЛЮЧЕНИЕ

В работе были предложены архитектура и метод обучения обобщающих нейросетевых классификаторов при малом размере обучающей выборки. Подход заключается в использовании древовидного классификатора и привлечении дополнительной общедоступной выборки большого объема. При этом происходит обучение двух классификаторов, имеющих общие входные (префиксные) слои. В отличие от опубликованных методов обучения переносом, при обобщающем обучении целевая и дополнительная выборка используются одновременно. Такой подход позволяет сократить переобучение на элементах целевой обучающей выборки за счет обнаружения и переноса на целевой классификатор информации об особенностях изображений, обнаруженных за счет обучающей выборки дополнительной задачи, и одновременно сохранить оптимальность префиксных слоев классификатора целевой задачи.

Проведенные численные эксперименты показали, что предложенный подход позволяет обучить более качественный классификатор, чем традиционный метод обучения переносом. Было показано, что для рассмотренных задач оптимальным является достаточно низкое значение относительной скорости обучения на целевой обучающей выборке. Эксперименты на различных дополнительных датасетах показали, что даже данные другой природы (рукописные цифры датасета MNIST) позволяют улучшить качество целевого (символы регистрационных номеров автомобилей) классификатора.

Максимальное качество целевого классификатора было достигнуто путем совмещения метода обобщающего классификатора и предварительной аугментации данных.

СПИСОК ЛИТЕРАТУРЫ

1. Krizhevsky Alex, Sutskever Ilya, Hinton Geoffrey E. Imagenet classification with deep convolutional neural networks // Advances in neural information processing systems. — 2012. — P. 1097–1105.
2. Multi-digit number recognition from street view imagery using deep convolutional neural networks / Ian J Goodfellow, Yaroslav Bulatov, Julian Ibarz et al. // arXiv preprint arXiv:1312.6082. — 2013.
3. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning / Hoo-Chang Shin, Holger R Roth, Mingchen Gao et al. // IEEE transactions on medical imaging. — 2016. — Vol. 35, no. 5. — P. 1285–1298.
4. Griffin Gregory, Holub Alex, Perona Pietro. Caltech-256 object category dataset. — 2007.
5. de Campos TE, Bodla RB, Varma M. The Chars74K dataset. — 2009.
6. Matching networks for one shot learning / Oriol Vinyals, Charles Blundell, Tim Lillicrap et al. // Advances in Neural Information Processing Systems. — 2016. — P. 3630–3638.
7. Koch Gregory, Zemel Richard, Salakhutdinov Ruslan. Siamese neural networks for one-shot image recognition // ICML Deep Learning Workshop. — Vol. 2. — 2015.
8. Parisotto Emilio, Ba Jimmy Lei, Salakhutdinov Ruslan. Actor-mimic: Deep multitask and transfer reinforcement learning // arXiv preprint arXiv:1511.06342. — 2015.
9. Massively Multitask Networks for Drug Discovery / Bharath Ramsundar, Steven Kearnes, Patrick Riley et al.
10. Devries Terrance, Biswaranjan Kumar, Taylor Graham W. Multi-task learning of facial landmarks and expression // Computer and Robot Vision (CRV), 2014 Canadian Conference on / IEEE. — 2014. — P. 98–103.
11. Tanner Martin A, Wong Wing Hung. The calculation of posterior distributions by data augmentation // Journal of the American statistical Association. — 1987. — Vol. 82, no. 398. — P. 528–540.
12. Povolotskiy Mikhail, Kuznetsova Elena, Khanipov Timur. Russian license plate segmentation based on dynamic time warping // European Council for Modeling and Simulation – ECMS 2017. — 2017. — P. 285–291.

13. LeCun Yann. The MNIST database of handwritten digits // http://yann. lecun. com/exdb/mnist/. — 1998.

14. Prechelt Lutz. Automatic early stopping using cross validation: quantifying the criteria. — Vol. 11. — Elsevier, 1998. — P. 761–767.

Development of architecture and training methods of multi-purpose neural networks for one-shot classification of images

Savchik A.V., Shvets E.A., Nikolaev D.P.

This paper considers a method for training a neural network image classifier using a small dataset by employing an additional (possibly unrelated) dataset of large size. We propose to train a multi-purpose artificial neural network that has a tree-like structure. First layers of the tree are shared between the classifiers for different tasks, but each leaf of the tree is an output layer of only one classifier. Full classifier for each task consists of a sequence of layers from the root of the tree to the corresponding leaf. We applied the proposed method to a task of one-shot recognition of symbols from vehicle license plates and compared the results of resulting classifier with results of classifiers trained by existing methods.

KEYWORDS: machine learning, artificial neural networks, multi-purpose learning, image classification.