Прикладные особенности обучения нейросетевых классификаторов в индустриальных задачах распознавания образов*

Алёна Иванова	Елена Кузнецова	Дмитрий Николаев
ИППИ РАН	ИППИ РАН	ИППИ РАН
abirisina@mail.ru	vojageur@gmail.com	dimonstr@iitp.ru

Аннотация. Описаны распространенные проблемы построения нейросетевых классификаторов на несбалансированных данных, полученных с сенсоров в режиме реального ограниченного времени. Предложен алгоритм синтеза данных с использованием известных методов обработки изображений для увеличения объема и устранения несбалансированности обучающей выборки. Приведены результаты вычислительных экспериментов, демонстрирующие повышение качества работы классификатора при использовании алгоритма синтеза данных на примере задачи классификации образов символов на фотографиях паспортов РФ. Рассмотрен вопрос построения векторов входных признаков классификатора на основе изображений обучающей выборки, предложен метод нормализации яркости изображений при формировании векторов признаков. Приведены вычислительные эксперименты, показывающие целесообразность использования регуляризации для улучшения обобщающей способности классификатора. Исследован вопрос выбора архитектуры классификатора, обеспечивающей наилучшее качество классификации при существующих ограничениях на быстродействие работы алгоритма в реальном времени.

Ключевые слова: машинное обучение, обучение на несбалансированных данных, синтез данных, нейронные сети, регуляризация, обработка изображений, компьютерное зрение, распознавание образов, распознавание символов, контрастирование изображений, бесшовная склейка изображений, сегментация документов.

---

* Работа частично финансово поддержана грантами РФФИ № 13-07-12178, № 13-07-12172.

1 Введение

1.1 Проблема построения классификатора с обучением на реальных данных

В настоящее время модули классификации на основе нейронных сетей являются наиболее эффективным инструментом для решения большого спектра лабораторных задач детектирования и классификации образов на изображениях [1-4]. Однако, в индустриальных задачах построение надежного классификатора затрудняется тем, что реальные исходные данные нередко бывают неточными в результате разного рода измерительных погрешностей, что усложняет их разделимость в пространстве входных признаков и может приводить к появлению т.н. «выбросов» (outliers) или даже противоречий (объекты со схожими наборами входных признаков относятся к разным классам) в признаковом пространстве. Следствием такого рода искажений, а также нередко характерной для реальных данных высокой вариативности объектов, принадлежащих одному классу (например, в задачах распознавания образов – изменений яркости, контрастности, пространственного расположения образов на изображении, присутствия шумов различной природы, теней, бликов, проективных искажений трехмерных объектов) является тенденция к неоднородности расположения данных в пространстве входных параметров, т.е. невозможность построения классификаторов, надежно отделяющих объекты каждого класса при малом числе свободных параметров обучаемого алгоритма. Устойчивость классификатора частично достигается путем усложнения структуры классификатора, однако, это неизбежно приводит к увеличению вычислительной сложности распознающего модуля, что может являться существенным недостатком при распознавании в режиме реального времени в условиях ограниченных вычислительных мощностей. Кроме того, известной проблемой в задачах обучения по прецедентам является тенденция к «переобучению» (overfitting) при выборе чрезмерно сложных моделей, обладающих избыточным числом свободных параметров, т.е. повышению точности работы на обучающей выборке с одновременным снижением обобщающей способности алгоритма. Т.о., существенным вопросом является выбор оптимальной архитектуры классификатора для эффективного решения поставленной задачи, а также построения входных признаков классификатора на основе исходных данных – с одной стороны, достаточно репрезентативных для дифференциации объектов различных классов и обеспечивающих отсутствие противоречий, с другой – обладающих максимальной нечувствительностью к различного рода искажениям объектов одного класса.

Распространенным способом ограничения сложности модели при большом числе свободных параметров является внесение регуляризации, т.е. добавление к минимизируемой функции ошибки слагаемого «штрафа» за чрезмерно сложные модели. При этом вопросы выбора минимизирующего функционала и параметра регуляризации (т.е. коэффициента при суммировании) также остаются открытыми для конкретной решаемой задачи.

Другой известной проблемой использования реальных данных для обучения машин является вопрос репрезентативности (т.е. достаточности для восстановления регрессии) имеющейся обучающей выборки. В общем случае, не существует неэмпирического алгоритма проверки гипотезы о репрезентативно-

сти имеющейся выборки. При этом исходные данные могут быть несбалансированными относительно реальных условий измерения или наблюдения, недостаточными для обучения классификатора (например, объектов существенно меньше, чем признаков), или, напротив, избыточными (обучение на слишком больших выборках может быть крайне медленным).

1.2 Постановка задачи

Для проведения вычислительных экспериментов по влиянию изменений архитектуры нейросетевого классификатора, параметров обучения и способа формирования векторов признаков из исходных данных на качество классификации используется задача классификации символов полей имени и фамилии на изображениях фотографий паспортов РФ. Входными данными для обучения являются одноканальные яркостные изображения символов кириллицы, а также разделяющих символов “-”, “ “, каждому изображению ставится в соответствие класс соответствующего символа. Кроме того, т.к. обучаемый классификатор входит в состав модуля автоматической сегментации печатных символов полей имени и фамилии паспортов, он должен характеризоваться способностью дифференцирования знакосимволов от т.н. класса ложных сегментов (изображений фрагментов символов, пар и троек различных символов, далее для обозначения класса ложных сегментов будет использоваться обозначение “~”).

Исходная выборка существенно не сбалансирована относительно различных классов знакосимволов (т.к. данные для обучения получены из реальных полей Имя-Фамилия изображений паспортов, где некоторые символы (например, “А”) встречаются существенно чаще других (например, “Ъ”)), число образцов для различных классов варьируется 5 до 2000.

В разделе 2 рассматриваются вопросы синтеза относительной репрезентативной относительно решаемой задачи и сбалансированной внутри каждого класса и относительно различных классов выборки для обучения классификатора на основе имеющихся данных.

В разделе 3 рассматривается вопрос выбора оптимального метода построения по исходным изображениям векторов входных признаков для классификатора.

Обучение производилось с использованием пакета расширения Neural Network Toolbox MATLAB [5], в качестве базовой архитектуры обучаемого алгоритма было принято использование нейронных сетей, состоящих из полносвязных слоев (fully-connected layers, FC-NN), в качестве функций активации использовалось преобразование tansig (гиперболический тангенс). В качестве метода обучения FC-NN использовался trainscg (метод шкалированных связанных градиентов) с минимизацией среднеквадратичной ошибки (minimum squared error, MSE) между целевыми векторами φ_{opt} и выходными векторами классификатора φ с регуляризирующим членом. Вопрос целесообразности использования регуляризации и выбора ее оптимальных параметров будет рассмотре в разделе 4.

В разделе 5 исследуется зависимость качества классификации от выбора архитектуры FC-NN, рассматривается вопрос выбора оптимальной архитектуры нейронной сети.

2 Синтез данных для обучения классификатора

Большинство стандартных алгоритмов машинного обучения, в т.ч. FC-NN, предполагают обучение на сбалансированных данных с равными стоимостями ошибки классификации для всех образцов обучающей выборки, что нередко является затруднительным в условиях обучения классификатора на основе данных, полученных при помощи каких-либо сенсоров в режиме онлайн, в условиях ограниченных временных и мощностных ресурсов, затрачиваемых на сбор данных для обучения. Проблемой обучения на несбалансированных данных является их способность значительно снизить качество обучения стандартных алгоритмов, т.к. они не обеспечивают требуемых характеристик распределения данных при обучении. Так, в [6, 7] приведен ряд вычислительных экспериментов с обучением классификаторов, показывающих снижение качества классификации при искусственной перебалансировке данных для обучения. Под несбалансированностью данных обычно понимается как неравномерность распределения данных между классами (например, в рассматриваемой задаче распознавания символов полей Имя-Фамилия, распределение символов, полученных на основе реальных изображений паспортов, неравномерно), так и неравномерность данных внутри класса, являющаяся следствием природы полученных данных (например, в распознающих видеосистемах реального времени с использованием алгоритмов машинного обучения, проблема неравномерности данных внутри классов связана с тем, что изображения видеоряда являются динамическими сценами, существенно меняющимися в зависимости от условий естественного и искусственного освещения, погодных условий, динамики фона распознаваемого образа, геометрии расположения объекта относительно регистрирующей камеры). Проблема обучения на несбалансированных данных является достаточно распространенной темой для исследований последних лет [8].

Существует два основных подхода к построению классификаторов на основе несбалансированных данных: построение сбалансированной выборки на основе исходной (data sampling) и модификация алгоритмов обучения на несбалансированных данных. К первому подходу относятся удаление/дублирование случайных элементов различных классов для достижения сбалансированности данных между классами (random undersampling/oversampling), обучение классификаторов на динамически формируемых на итерациях обучения классификатора сбалансированных подмножествах исходной выборки (informed undersampling [9], предварительная кластеризация исходных данных с последующим выравниванием объема кластеров схожих элементов исходной выборки для устранения внутренней и межклассовой несбалансированности (cluser-based sampling method) [10, 11] синтез векторов входных признаков для обучения классификатора (искусственные вектора признаков формируются как произведения векторов признаков случайных близких по выбранной метрике сходства в пространстве исходных данных образцов), кластеризация элементов каждого класса исходной выборки с выбранной метрикой близости с последующим удалением выбросов (sampling with data cleaning) [12]. Ко второму подходу относятся алгоритмы с модификацией матрицы стоимости ошибки классификации для различных классов/кластеров несбалансированной выборки (cost-sensitive learning) [13, 14].

Целью данной работы является обучение FC-NN на несбалансированных данных без модификации алгоритма обучения, т.о. далее авторами рассматривается только первый подход. Т.к. в рассматриваемой задаче в исходной выборке для обучения образцов редко встречающихся в рассматриваемой предметной области символов недостаточно для обучения достаточно надежного классификатора (число образцов для различных классов варьируется от 5 до 2000, при этом класс ложных сегментов в исходной выборке отсутствует полностью), авторами использовался подход с расширением обучающей выборки путем синтеза данных. Известен ряд специфических для задачи распознавания образов символов методов расширения обучающей выборки для распознавания рукописных текстов на основе упругих деформаций, эмулирующих колебания мышц руки в процессе написания текста [15, 16] , а также локальных и глобальных афинных искажений изображений [15, 17,18].

В контексте решаемой задачи обучения классификатора для модуля автоматической сегментации символов полей Имя-Фамилия фотоизображений паспортов (рис 1 a), требуется сформировать обучающую выборку для построения векторов признаков, максимально приближенным к условиям наблюдения, на основе отсканированных изображений документов (рис 1 b). Вектора признаков для обучения классификатора на основе исходных изображений формировались как последовательность пикселей строк изображений, приведенных к единому размеру (вопрос выбора оптимального размера изображений будет рассмотрен далее). Для моделирования изображений, схожих с реальными данными, использовались следующие методы синтеза:

1. Моделирование перспективных искажений, возникающих при фотосъемке: афинные преобразования изображений.
2. Моделирование эффектов фотооцифровки изображения: размытие изображений с ядром Гаусса, гамма-коррекция, “стирание” локальных областей символа на изображении, случайное зашумление изображения.
3. Моделирование эффектов изменения освещения при фотосьемке: моделирование теней, изменение глобального контраста изображения.

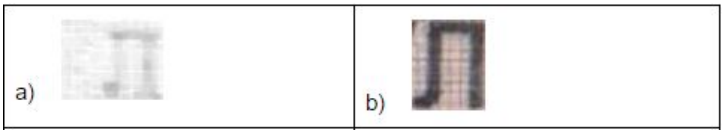

Рис. 1. Примеры образцов реальных данных и обучающей выборки.

Кроме того, в результате работы модуля предварительной сегментации символов возникает значимая доля ошибок сегментации, т.н. “ложных сегментов”, не являющихся изображениями символов (пар, троек символов, фрагментов символов). Выделенные таким образом изображения подаются на вход классификатора, следовательно требовалось обеспечить надежное распознавание таких ложных сегментов на уровне классификатора.

Экспериментально было выявлено, что присутствующие в исходной обучающей выборке сочетания символов не содержат достаточного числа образцов определенных сочетаний символов/их фрагментов, соответствующих характер-

ным ошибкам обученного классификатора (примеры типичных ошибок классификатора представлены на рис. 2). Поэтому для варьирования размера и содержания образцов класса “~” в обучающей выборке использовались искусственно “склеенные” случайно выбранные изображения определенных символов/их фрагментов, при этом доля образцов сочетаний каждого фиксированного набора символов/фрагментов регулировалась в соответствии с частотой возникновения соответствующих ошибок классификации при тестировании на реальных данных.

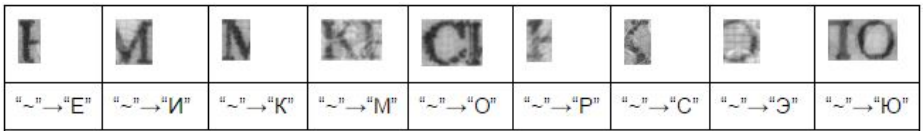

Рис. 2. Примеры характерных ошибок классификатора на ложных сегментах.

На рис. 3 a-c видно существенное различие фонов различных символов, что порождает видимые границы наложения в результате их “склейки” (рис. 3 b), существенно зашумляющие вектора входных признаков при добавлении таких образцов к классу “~”. Для “склеивания” изображений без образования видимых артефактов на границах наложения изображений был использован метод бесшовной склейки с применением преобразования Пуассона [19] (результат представлен на рис. 3 c).

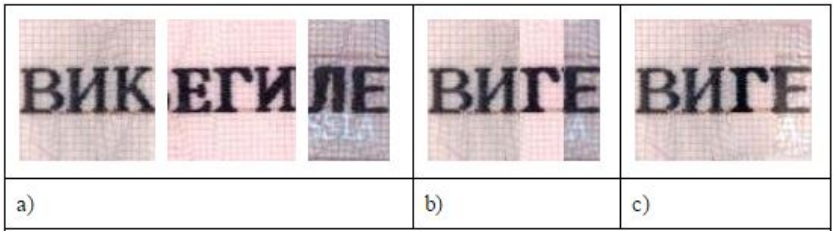

Рис. 3. Склеивание фрагментов изображений для генерации класса “~”: a) исходные фрагменты изображений; b) результат “наивной” склейки изображений; c) результат “бесшовной” склейки изображений.

Кроме того, классификатор, должен обеспечивать робастность к неточностям в работе модуля предварительной сегментации (наличие фоновой рамки переменного размера вокруг символа, поворот символа на изображении соответствующего сегмента). Для моделирования неточностей сегментации к исходным (размеченным оператором) границам полей символов на изображении применялись случайные преобразования сдвига каждого из полей, к изображениям символов - преобразования поворота на случайный угол.

Синтез каждого образца для расширения обучающей выборки осуществлялся путем применения к случайно выбранному изображению исходной выборки случайного поднабора из описанных выше преобразований. При этом вероятность применения каждого преобразования, а также параметры самих преобразований являлись параметрами синтеза данных и выбирались экспериментально

как соответствующие минимуму ошибки классификации (на тестовых данных) при обучении на синтезированных данных. В таблице 1 приведены параметры синтеза данных, соответствующие каждому из использованных преобразований исходных образцов.

Table 1[1]**.** Параметры синтеза данных при расширении исходной обучающей выборки.

1. Перспективное преобразование - смещение каждой из четырех угловых точек исходного окаймляющего прямоугольника символа не более чем на x_{rate}/y_{rate} от ширины/высоты прямоугольника по горизонтали/вертикали соответственно. Доли сдвигов $d_x \in [-x_{rate}, x_{rate}], \sigma n$; $d_y \in [-y_{rate}, y_{rate}], \sigma n$.
Изображение преобразуется с матрицей проективного преобразования, переносящей исходные точки окаймляющего прямоугольника в смещенные.

Рис. 4. Примеры случайных перспективных преобразований изображения с x_{rate}=0.08, y_{rate}=0.08.

2. Размытие изображения - сглаживание с ядром Гаусса с параметром $\sigma \in [\sigma_{min}, \sigma_{max}], \sigma n$.

Рис. 5. Примеры случайных преобразований размытием, где σ = 0, 0.5, 1.0, 1.4.

3. Гамма-коррекция - применение с вероятностью к σu к исходному яркостному изображению гамма-коррекции со случайно сгенерированным параметром $p \in [p_{min}, p_{max}]$.

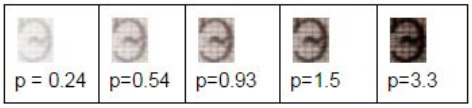

Рис. 6. Примеры случайных преобразований гамма-коррекцией.

4. Зашумление изображения - для каждого пиксела с вероятностью pu $i'_{x, y} = \max(0, (\min(1, i_{x, y} + di(x, y)))$, $di(x,y) \in [0, di] \in [di_{min}, di_{max}]$.

Рис. 7**.** Примеры случайных преобразований зашумлением, где di = 0.07, 0.2, 0.4.

---

[1]Список условных обозначений, использованных в таб. 1:

ix, y - значение яркости точки с координатами (x, y) исходного изображения

i'x, y - значение яркости точки с координатами (x, y) результирующего изображения

a $\in$ [a, b] - значение a выбирается как случайная равномерно распределенная величина на отрезке [a, b].

a $\in$ [a, b], σ - значение a выбирается как случайная нормально распределенная с параметром σ величина на отрезке [a, b].

5. “Стирание” локальных областей символа на изображении -
$i'_{x,y} = b_{x,y} * j_{x,y} + (1 - b_{x,y}) * i_{x,y,}$ где $b_{x,y} = \min(1, m_{x,y})$, где
$m_{x,y}$ - значения яркости соответствующих пикселов на изображении-маске m. m формируется как размытое фильтром Гаусса ($\epsilon [\sigma_{min}, \sigma_{max}]$, σn) изображение, заполненное нулями и случайными значениями nzxy $\epsilon [0, nz_{max}]$, σn_{nz}.

Рис. 8. Примеры случайных преобразований замещением локальных областей изображения фоном с $\sigma_{min}=0.5$, $\sigma_{max}=3.5$, $nz_{max}=2$, $\sigma n_{nz}=0.3$.

6. Моделирование теней - с вероятностью p изображение делится на две части случайно сгенерированной прямой, для одной из частей
$i'_{x,y} = k * i_{x,y}$, где $k \epsilon [k_{min}, k_{max}]$, $0 < k_{min} <= k_{max} < 1$.

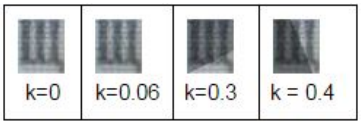

Рис. 9. Примеры случайных преобразований с моделированием теней.

7. Добавление фона вокруг образца - случайный сдвиг каждой стороны исходного окаймляющего прямоугольника символа не более чем на доли x_{rate}/y_{rate} от ширины/высоты исходной рамки, соответственно. Доли сдвигов
$dx \epsilon [-x_{rate}, x_{rate}]$, σn_x; $dy \epsilon [-y_{rate}, y_{rate}]$, σn_y.

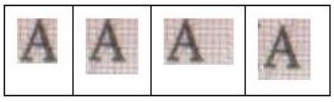

Рис. 10. Примеры случайных преобразований с добавлениями фона к изображению образца с $x_{rate}=1$, $y_{rate}=0.3$, $\sigma n_x=5.0$, $\sigma n_y=7.0$.

8. Сдвиг окаймляющей рамки образца - с вероятностью p сдвиг по горизонтали на долю от ширины изображения $dx \epsilon [-x_{rate}, x_{rate}]$, σn_x, по вертикали - на долю от высоты $dy \epsilon [-y_{rate}, y_{rate}]$, σn_y.

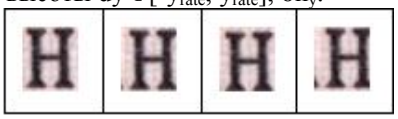

Рис. 11. Примеры случайный преобразований со сдвигом рамки образца с $x_{rate}=0.05$, $y_{rate}=0.05$.

9. Поворот изображения - угол поворота $\alpha \epsilon [-\alpha, \alpha]$, σn.

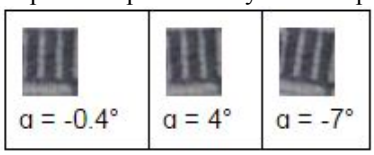

Рис. 12. Примеры случайных преобразований поворотом изображения.

В таблице 2 приведены результаты ряда вычислительных экспериментов с обучением классификаторов на расширенных обучающих выборках, полученных путем синтеза данных с различными параметрами.

Table 2[2]. Зависимость качества работы классификатора от параметров синтеза данных.

Эксперимент			Random over-sampling (№1)	Синтез данных №2	№3	№4	№5	№6	№7	№8	№9	№10	№11
Качество классификации на тестовой выборке (суммарное число ошибок)			66% (176)	63% (193)	52% (250)	55% (237)	60% (210)	63% (195)	68% (170)	61% (202)	62% (200)	55% (237)	64% (190)
Параметры синтеза	1. Песпективное преобразование	σn_x	—	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
		σn_y	—	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
		x_{rate}	—	0,08	0,20	0,40	0,08	0,08	0,08	0,08	0,08	0,08	0,08
		y_{rate}	—	0,08	0,20	0,40	0,08	0,08	0,08	0,08	0,08	0,08	0,08
	2. Сглаживание изображения	σ_{min}	—	0,00	0,70	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,50
		σ_{max}	—	1,50	1,50	1,80	1,50	1,50	1,50	1,50	1,50	1,50	3,50
		σn	—	1,20	1,70	1,20	1,20	1,20	1,20	1,20	1,20	1,20	4,20
	3. Гамма-коррекция	σu	—	0,20	0,20	1,20	0,80	0,20	0,20	0,20	0,20	0,20	0,80
		p_{min}	—	0,25	0,55	0,50	0,90	0,25	0,25	0,25	0,25	0,25	0,50
		p_{max}	—	1,50	1,80	1,10	2,50	1,50	1,50	1,50	1,50	1,50	3,50
	4. Случайный шум	pu	—	0,3	0,5	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,2
		di_{min}	—	0,04	0,1	0,1	0,04	0,04	0,04	0,04	0,04	0,04	0,04
		di_{max}	—	0,2	0,3	0,3	0,2	0,2	0,2	0,2	0,2	0,2	0,2
	5.“Стирание” локальных областей символа	σ_{min}	—	0,50	0,40	0,50	0,50	0,50	0,50	0,50	0,10	0,80	0,50
		σ_{max}	—	3,50	2,50	3,50	3,50	3,50	3,50	3,50	1,00	5,00	3,50
		nz_{max}	—	2,00	7,00	2,00	3,00	2,00	2,00	2,00	1,00	10,0	5,00
		σn	—	0,40	0,40	0,30	0,60	0,40	0,40	0,40	0,20	0,70	1,00
		σn_{nz}	—	0,30	0,30	0,70	19,0	0,30	0,30	0,30	0,50	0,80	9,00
	6. Синтез теней	p	—	0,05	0,05	0,10	0,05	0,05	0,05	0,05	0,05	0,05	0,00
		k_{min}	—	0,10	0,10	0,10	0,10	0,10	0,10	0,10	0,10	0,10	0,10
		k_{max}	—	0,40	0,40	0,40	0,40	0,40	0,40	0,40	0,40	0,40	0,40

[2] В данной серии вычислительных экспериментов в качестве входных векторов признаков использовалась последовательность строк изображений, приведенных к единому размеру 11*11 и нормализованных по яркости автоконтрастированием (см. раздел 3), обучение осуществлялось с использованием регуляризации (см. раздел 4), использовалась архитектура FC-NN с двумя внутренними

полносвязными слоями с числом нейронов 265 и 64 на 1 и 2 слоях соответственно (вопросу выбора оптимальной архитектуры FC-NN посвящен раздел 5)

	7. Расширение окаймляющей рамки образца	σn_x	—	0,80	0,50	0,80	0,80	0,80	0,80	0,80	0,80	0,80	1,80
		σn_y	—	0,80	0,50	0,80	0,80	0,80	0,80	0,80	0,80	0,80	1,80
		x_{rate}	—	0,05	0,40	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,20
		y_{rate}	—	0,05	0,40	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,20
	8. Сдвиг рамки образца	p	—	0,40	0,70	0,70	0,40	0,10	0,80	1,00	0,40	0,40	0,80
		σn_x	—	6,00	3,00	6,00	6,00	2,00	10,0	14,0	6,00	6,00	10,0
		σn_y	—	6,80	1,80	6,80	6,80	2,00	10,0	14,0	6,80	6,80	10,8
	9. Поворот изображения	α	—	7,00	7,00	15,0	7,00	7,00	7,00	7,00	7,00	7,00	7,00
		σn	—	1,30	1,30	1,30	1,30	1,30	1,30	1,30	1,30	1,30	1,30
10. Объём обучающей выборки			5000	5000	2000	2000	15000	5000	5000	5000	5000	5000	25000
11. Отношение числа образцов ложных сегмен-тов к числу обычных			3,0	3,0	3,0	1,0	3,0	3,0	3,0	3,0	3,0	3,0	3,0

Из таблицы 2 видно, что расширение обучающей выборки на основе предложенного алгоритма синтеза данных позволяет варьировать характеристики обучающей выборки (эксперименты №№ 2-11), приводит к повышению качества работы классификатора в сравнении с простейшим подходом к выравниванию межклассовой несбалансированности random oversampling (описание см. выше) при равных объемах обучающей выборки (см. эксперименты № 1 и № 7). По результатам серии экспериментов (№№ 2-11) с различными параметрами синтеза данных и объемами обучающих выборок, был выбран оптимальный (№ 7), превосходящий по качеству классификатор, полученный на выборке с random oversampling в на 2% или в 1,03 раза. Тестировались обученные на полученных данных классификаторы на выборке, содержащей 524 элемента. Полученная на данном этапе оптимальная обучающая выборка будет использована в дальнейших экспериментах с выбором оптимальных векторов признаков, архитектуры и параметров обучения классификатора (см. разделы 3-5)

Кроме того, предложенный метод синтеза данных позволяет осуществлять итеративное улучшение качества обучения классификаторов за счет дополнения обучающей выборки кластерами образцов, схожих с характерными ошибками классификатора, обученного на предыдущей итерации.

3 Нормализация исходных данных и построение векторов входных признаков для обучения классификатора

Вектора входных признаков для обучения классификатора формировались в виде последовательности яркостных значений пикселов изображений (значения

яркости - 32-битные числа с плавающей точкой в интервале [0, 1]), упорядоченных в порядке возрастания пар координат (y, x) на исходных изображениях обучающей выборки, предварительно приведенных к единому размеру. При этом существенным является вопрос выбора размеров исходного изображения, т.к., с одной стороны, использование необоснованно больших изображений повышает вычислительную сложность обучаемого классификатора, а также приводит к “замусоренности” входных векторов признаков несущественными или избыточными значениями, отражающими шумы различной природы на фотоизображениях, с другой - слишком большое сжатие изображений может привести к утере существенной информации об образах объектов, вследствие - снижению качества классификации обученной машины.

Выбор оптимального для данной задачи размера вектора входных признаков производился экспериментально. В таблице 3 приведены результаты ряда вычислительных экспериментов с обучением классификатора на векторах признаков, полученных на основе изображений, приведенных к различным размерам. Здесь и далее в качестве исходной обучающей выборки для последующих будут использованы результаты, соответствующие наиболее успешному из последних экспериментов.

Table 3. Зависимость качества классификатора от размера вектора признаков.

Эксперимент	Размер вектора признаков	Автоконтраст	Качество (суммарное число ошибок)
№ 1	7*7	без автоконтраста	51% (255)
№ 2	11*11	без автоконтраста	55% (234)
№ 3	20*20	без автоконтраста	54% (242)

По результатам экспериментов, в качестве оптимального размера растра для формирования векторов входных признаков первого слоя классификатора был выбран размер 11*11 (эксперимент №2).

Стандартным приемом предобработки входных данных для обучения классификаторов является нормализация (т.е. приведение к одному диапазону) разнородных признаков, измеряемых в разных величинах и принимающих значения в различных диапазонах (напр., температура, плотность, и т.д.). Такая нормализация признаков позволяет ускорить процесс обучения классификатора, а также повысить качество работы классификатора [20].

В рассматриваемой задаче входные данные однородны и соответствуют яркостным значениям пикселей изображения в диапазоне от 0 до 1, однако, абсолютные значения яркости соответствующих пикселей изображений одного класса могут существенно варьироваться ввиду изменений контраста изображения в результате изменений освещения при съемке изображений. В [21, 22] для улучшения цветового расширения входных изображений применяется автоконтрастирование (для используемого формата входных данных - растяжение шкалы яркости изображения от 0 до 1). Однако, в рассматриваемой задаче подобное автоконтрастирование может привести к усилению несущественного “шума” на входных изображениях. Так, на рис. 13 c, 13 d представлен результат автоконтрастирования яркости изображений, соответствующих классам “ “ (рис. 13 a) и

“~” (рис. 13 b). Т.е. негативным эффектом данного способа предобработки изображений может стать принципиальная неразделимость классов “ “-“~”, и, как следствие, ухудшение качества разбиения элементов между данными двумя классами классификатором.

Для снижения влияния эффекта усиления шума в результате автоконтрастирования, авторами предлагается использование модифицированного алгоритма автоконтрастирования (т.н. “умный” автоконтраст) изображения:

$i_{hgh} \rightarrow 0$

$i_{low} \rightarrow \min(1, k * (i_{hgh} - i_{low}))$, где

i_{low}, i_{hgh} - значения яркости, соответствующие квантилям q_{low} и q_{hgh} соответственно на гистограмме яркости изображения. Результат описанной модификации автоконтрастирования представлен на рис. 13 e, 13 f.

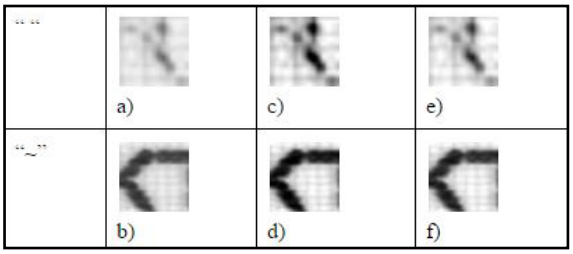

Рис. 13. Результат применения различных видов автоконтрастирования: a,b) исходное изображение; c,d) обычный автоконтраст; e,f) "умный" автоконтраст.

Результаты вычислительных экспериментов без нормализации яркости, с использованием стандартного автоконтрастирования, а также описанной выше модификации с различными параметрами k, q_{low}, q_{hgh} представлены в таблице 4.

Table 4. Результаты применения различных видов автоконтрастирования.

Эксперимент	Предобработка растров при формировании векторов признаков	Качество работы обученного классификатора (суммарное число ошибок)
№ 1	без автоконтрастирования	55% (234)
№ 2	автоконтрастирование	22% (411)
№ 3	“умное” автоконтрастирование, k = 2	60% (211)
№ 4	“умное” автоконтрастирование, k = 3	63% (194)

Из таблицы 4 видно, что использование обычного автоконтрастирования растров при формировании векторов признаков приводит к существенному снижению качества работы обученного классификатора, главным образом за счет существенного роста ошибок классификации на классах “~” и “ ”. Однако, предложенный алгоритм “умного” автоконтрастирования позволяет повысить качество работы обученного классификатора (эксперименты №№3-4). В качестве оптимального метода предобработки изображений при построении векторов признаков было выбрано “умное” автоконтрастирование с k = 3 (см. экспери-

мент №4), позволившее повысить качество работы классификатора в 1,15 раз в сравнении с экспериментом без использования автоконтрастирования (№1)

4 Выбор параметров обучения нейронной сети

Существенной проблемой обучения машин является проблема мультиколлинеарности, связанной с плохой обусловленностью матрицы ковариации входных параметров ввиду сильной корреляции или линейной зависимости части входных данных. В таком случае результат сходимости итерационного процесса обучения классификатора существенно зависит от присутствия даже малых погрешностей измерения входных параметров, существенно снижается обобщающая способность обучаемого алгоритма. Другая причина переобучения связана с выбором чрезмерно сложной модели обучаемого классификатора со слишком большим числом свободных параметров - при обучении классификатор “затачивается” под особенности обучающих данных, при этом снижается обобщающая способность.

Распространенным методом борьбы с переобучением является добавление к оптимизируемому при обучении функционалу регуляризатора. Для FC-NN показано, что уменьшение весов связей позволяет повысить ее обобщающую способность [23]. Для сокращения весов к минимизирующему функционалу квадратичной ошибки (squared error, se) добавляется штрафное слагаемое

$$Q' = se + \tau/2 * ||w||^2 \quad \text{, где}$$

τ - параметр регуляризации, варьирование которого при обучении позволяет найти компромисс между точностью настройки весов на обучающей выборке и устойчивостью весов.

В таблице 5 приведен ряд вычислительных экспериментов с различными значениями параметра регуляризации.

Table 5. Качество классификатора с применением регуляризации и без нее.

Эксперимент	Качество работы классификатора (суммарное число ошибок)	
	Без регуляризации (значение параметра регуляризации = 0)	С регуляризацией (значение параметра регуляризации = 0,05)
№ 1	48% (273)	52% (250)
№ 2	53% (248)	55% (237)
№ 3	61% (204)	60% (210)
№ 4	60% (211)	63% (195)

Результаты экспериментов показывают, что добавление регуляризирующего члена к минимизирующему функционалу при обучении классификатора позволило повысить качество работы полученного классификатора в 1,1 раз.

5 Выбор архитектуры нейронной сети

Как правило выбор оптимальной архитектуры FC-NN осуществляется эмпирически для конкретной решаемой задачи и имеющегося набора данных для обучения. Следствием выбора модели с чрезмерно малым числом свободных коэффициентов является большое значение функции ошибки при обучении классификатора, недостаточная точность классификации на тестовых данных. Увеличение числа свободных параметров модели повышает вычислительную сложность алгоритма классификации, чрезмерно сложная модель имеет тенденцию к переобучению.

В таблице 6 приведены результаты ряда вычислительных экспериментов по обучению классификаторов с различным числом полносвязных слоев и числом нейронов на каждом слое.

Результаты экспериментов показали, что усложнение архитектуры классификатора не гарантирует роста его качества на тестовых данных (напр., см. эксперименты № 1, 7). Кроме того, в таблице 6 можно заметить тенденцию роста качества работы классификаторов при сокращении числа внутренних слоев классификатора. Данная закономерность была выявлена эвристически для рассматриваемой задачи, в общем случае требование к присутствию этой закономерности не обосновано. в качестве оптимальной архитектуры по результатам экспериментов была выбрана 2-слойная архитектура с 256-ю нейронами на 1-м и 64-мя нейронами на 2-м слоях, соответственно (эксперимент № 1).

Table 6. Результаты обучения классификатора с разными типами архитектур.

Эксперимент	Число слоев	Число нейронов на каждом слое				Вычислительная сложность	Качество работы классификатора (суммарное число ошибок)
№ 1	2	256	64			49 664	68% (170)
№ 2		350	120			88 670	66% (177)
№ 3		600	250			231 600	69% (165)
№ 4	3	144	72	36		31 680	62% (200)
№ 5		200	88	45		47 380	64% (190)
№ 6		130	106	83		41 296	61% (204)
№ 7		288	144	72		89 280	63% (196)
№ 8		512	256	64		211 712	68% (169)
№ 9		1024	256	64		404 736	69% (163)
№ 10		256	1024	64		360 960	68% (167)
№ 11	4	100	100	100	100	45 700	62% (199)
№ 12		130	106	83	62	45 686	61% (204)
№ 13		220	160	110	60	88 180	67% (174)
№ 14		400	300	200	100	252 000	66% (180)
№ 15		875	300	100	25	401 775	68% (169)
№ 16		875	300	100	36	403 271	68% (167)

Тем не менее, сохраняется устойчивый порог некоторого числа ошибок, что можно объяснить тем, что исходная обучающая выборка плохо аппроксимирует тестовую. Причем во всех экспериментах преобладающее большинство ошибок

относилось к одному классу, а именно к ложным сегментам, где содержится один символ с широким окаймляющим полем (ложноположительная классификация). Так же ошибки группировались на символах с чуть менее широким полем, смещенным относительно центра рамки (ложноотрицательная классификация).

Таким образом, причина устойчивости возникновения ошибок данного типа заключается в сложности однозначно отделить такого рода ложные сегменты от верных, и наличии противоречий в тестовых данных, часто возникающих по этой причине на практике.

В качестве дальнейшего решения данной проблемы предлагается модифицировать структуру классификатора, путём добавления "промежуточного" класса для каждого символа, в рамках которого классификация изображения как данного символа и классификация как ложного сегмента не будет считаться ошибкой.

6 Заключение

В данной работе показано, что расширение обучающей выборки с использованием синтеза данных позволяет повысить качество классификации при работе с несбалансированными данными для обучения. Эффективность предложенного алгоритма продемонстрирована на примере построения нейросетевого классификатора для решения задачи распознавания символов полей имя-фамилия, полученных с фотографий паспортов РФ. Также предложен подход к нормализации данных при построении входных векторов признаков классификатора на основе изображений, приведены эксперименты, показывающие его эффективность для данной задачи.

Была экспериментально продемонстрирована целесообразность использования эмпирических методов при решении вопросов выбора архитектуры и параметров обучения нейросетевых классификаторов. Показана эффективность использования регуляризации для улучшения обобщающей способности классификатора. Исследован вопрос выбора оптимальной архитектуры классификатора, показано, что усложнение архитектуры не гарантирует роста качества классификации.

7 Литература

1. D. Cirean, etc. Multi-column deep neural networks for image classification. IEEE Conference on Computer Vision and Pattern Recognition, 2012.
2. D. Cirean, etc. Flexible, High Performance Convolutional Neural Networks for Image Classification. Proceedings of the Twenty-Second international joint conference on Artificial Intelligence, V. 2, pp. 1237-1242, 2013.
3. O. Russakovsky et al.. ImageNet Large Scale. Visual Recognition Challenge, 2014.
4. The Face Detection Algorithm Set To Revolutionize Image Search. Technology Review, February 16, 2015.
5. MathWorks Matlab Documentation official page: http://www.mathworks.com/help/matlab/

6. Gary M. Weiss , Forest Provost. The Effect of Class Distribution on Classifier Learning: An Empirical Study. Journal of Artifical Intelligence Research, 2003.
7. A. Estabrooks, T. Jo, N. Jakpowicz. A Multiple Resampling Method for Learning from Imbalanced Data Sets. Computational Intelligence, V. 20, Is. 1, 2004.
8. Haibo He, Edwardo A. Garcia. Learning from Imbalanced Data. IEEE Transactions on Knowledge and Data Engineering, pp. 1263-1284, September, 2009.
9. X.Y. Liu, J. Wu, and Z.H. Zhou. Exploratory Under Sampling for Class Imbalance Systems, Man, and Cybernetics, Part B: Cybernetics. IEEE Trans V. 39, Is. 2, 2008.
10. T. Jo and N. Japkowicz. Class Imbalances versus Small Disjuncts. ACM SIGKDD Explorations Newsletter - Special issue on learning from imbalanced datasers, V. 6, Is. 1, pp. 40-49, New York, USA, June 2004.
11. B.X. Wang and N. Japkowicz. Imbalanced Data Set Learning with Synthetic Samples. Proc. IRIS Machine Learning Workshop, Ottawa, 2004.
12. G.E.A.P.A. Batista, R.C. Prati, and M.C. Monard. A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data. ACM SIGKDD Explorations Newsletter - Special issue on learning from imbalanced datasers, V. 6, Issue 1, pp. 20-29, June 2004.
13. P. Domingos. MetaCost: A General Method for Making Classifiers Cost-Sensitive. 5 ACM SIGKDD International Conference on Knowledge discovery and data mining, p. 155-164, New York, 1999.
14. X.Y. Liu and Z.H. Zhou. Training Cost-Sensitive Neural Networks with Methods Addressing the Class Imbalance Problem. IEEE Trans. Knowledge and Data Eng., 2006.
15. Dan Claudiu Ciresan, Ueli Meier, Luca Maria Gambardella, Jurgen Schmidhuber. Deep Big Simple Neural Nets Excel on Hand-written Digit Recognition. Neural Computation, V. 22, №12, 2010.
16. Patrice Y. Simard, Dave Steinkraus, John C. Platt. Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis. ICDAR '03 Proceedings of the Seventh International Conference on Document Analysis and Recognition, 2003.
17. Larry Yaeger, Richard Lyon, Brandyn Webb. Effective Training of a Neural Network Character Classifier for Word Recognition. Advances in Neural Information Processing Systems, V. 9, Cambridge, 1997.
18. Жуковский А.Е., Тарасова Н.А., Усилин С.А., Николаев Д.П. Синтез обучающей выборки на основе реальных данных в задачах распознавания изображений. Информационные технологии и системы (ИТиС'12): сборник трудов конференции. М., 2012. С. 377-382.
19. P. Perez, M. Gangnet, A. Blanke. Poisson Image Editing. 03 ACM SIGGRAPH 2003 Papers, p.p. 313-318, 2003.
20. J.Sola, J. Sevilla. Importance of input data normalization for the application of neural networks to complex industrial problems. IEEE Trans. Nuclear Science, V. 44, № 3, p.p. 1464– 1468, 1997.
21. M. Martins, etc. A new method for multi-texture segmentation using neural networks. Neural Networks. Processin of the 2002 International Join Conference, V. 3, 2002.
22. S. Mato, Y. Yadav. Effectual Approach for Facial Expression Recognition System. International Journal of Advanced Research in Compiter and Communication Engineering, V. 4, May 2015.
23. Krogh A., Hertz J.A. A simple weight decay can improve generalization. Advances in Neural Informatio Processing Systems 4, pp. 950-957, 1992.