

Зависимость точности и оценки распознавания от степени различия между базами обучения и распознавания

М. Б. Гавриков, Н. В. Пестрякова,
А. В. Усков, В. В. Фарсобина

Исследована зависимость точности и оценки распознавания от степени различия между базами обучения и распознавания. Изучение проведено для баз распознавания, которые получены при помощи четырех моделей «разрушения» обучающей базы.

Введение

Интуитивно понятно, что любая база распознавания — это каким-либо образом «испорченная» база обучения. На практике бывает трудно определить, существует ли некая закономерность преобразования базы обучения в базу распознавания. В настоящей работе мы сами предлагаем четыре модели «разрушения» базы обучения, и в рамках каждой из них исследуем зависимость качества распознавания от степени «искажения» обучающей базы.

Будем анализировать поведение таких характеристик качества распознавания символов, как оценка распознавания (средняя по всем верно распознанным растрам базы), а также число неправильно распознанных образов по базе.

Скорость (абсолютная, средняя) между двумя точками на графике равна модулю изменения величины, откладываемой по оси ординат, отнесенному к модулю изменения величины, откладываемой по оси абсцисс.

Метод распознавания описан в работах [1–6].

1. Модель «наихудшего» разрушения

В обучающей базе на этапе распознавания для каждого растра выбирается случайным образом определенное количество пикселей, в которых v_i , лежащие в полуинтервале $0 \leq v_i < 0,5$, заменяются значением 1: $[0; 0,5) \rightarrow 1$, а при $0,5 \leq v_i \leq 1$ величиной 0: $[0,5; 1] \rightarrow 0$. Число пикселей, в которых производятся такие модификации, варьируется от 1 до некоторого значения, в данном случае 10, при котором преодолевается «порог» количества неправильно распознанных растров (мы устанавливаем его равным удвоенному числу изображений, нераспознанных на самой базе обучения, а именно, $881 \times 2 = 1762$). При каждом фиксированном количестве k ($1 \leq k \leq 10$) заменяемых пикселей осуществляется распознавание по всей базе. При построении графиков формально добавляется значение числа измененных пикселей, равное 0, что соответствует распознаванию обучающей базы.

2. Модель «случайного» разрушения

В обучающей базе на этапе распознавания для каждого растра выбирается случайным образом определенное количество пикселей, в которых v_i , лежащее на отрезке $0 \leq v_i \leq 1$, заменяется на случайное значение, также принадлежащее отрезку $[0, 1]$. Число пикселей, в которых производятся такие модификации, варьируется от 1 до 256, поскольку в данном случае не преодолевается «порог» количества неправильно распознанных растров ($881 \times 2 = 1762$).

Здесь и выше используется генератор псевдослучайных чисел. Так в каждом текущем растре для случайного выбора k различных пикселей из имеющихся перенумерованных от 0 до 255 (т. е. всего 256 пикселей), выполняются следующие действия. Для натурального числа, предлагаемого генератором псевдослучайных чисел, находим остаток от деления на 256. Он является целым и принадлежит требуемому отрезку $[0, 255]$. Дополнительно проводится проверка отличия нового номера от полученных прежде.

Аналогично, при выборе случайного значения v_i , принадлежащего отрезку $[0, 1]$, для натурального числа, предлагаемого генератором псевдослучайных чисел, находим остаток от деления на 101. Он является целым и принадлежит отрезку $[0, 100]$. Делим его на 100. Результат принадлежит отрезку $[0, 1]$.

На рис. 1а, б для сравнения приводятся результаты, полученные при «наихудшем» и «случайном» разрушении для числа пикселей, в которых

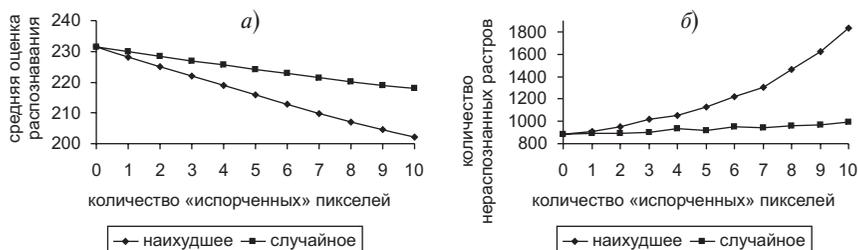


Рис. 1

производятся такие модификации, изменяющемся от 1 до 10. Также добавлен 0, что соответствует отсутствию модификаций.

На рис. 1а на отрезке для числа «испорченных» пикселей $[0, 10]$ показано линейное падение средней оценки распознавания для «наихудшего» разрушения; для «случайного» разрушения убывание почти линейное, но средняя скорость его приблизительно в два раза меньше. В обоих случаях зависимости являются строго монотонными.

На рис. 1б можно видеть монотонное нарастание количества неправильно распознанных растров при увеличении числа «испорченных» «наихудшем» образом пикселей. Когда их количество становится равным 10, число неправильно распознанных растров преодолевает «пороговое» значение и становится равным 1835. Для «случайного» разрушения график является почти монотонно нарастающим (наблюдаются небольшие колебания) и близким к линейному. При 10 модифицированных пикселях количество нераспознанных растров увеличивается незначительно, до значения 994. То есть на рассматриваемом отрезке для числа «испорченных» пикселей $[0, 10]$ средняя скорость увеличения числа неправильно распознанных растров для «наихудшего» разрушения более чем в семьдесят три раза превышает соответствующий показатель при «случайном» разрушении.

Как уже было отмечено ранее, при «случайном» разрушении пикселей «пороговое» количество нераспознанных растров не достигается даже при модификации всех пикселей. На рис. 2а видно, что средняя оценка распознавания монотонно убывает на максимально возможном отрезке для числа «испорченных» пикселей $[0, 256]$, но скорость падения этой величины на участке $[16, 40]$ существенно замедляется и далее остается очень незначительной. Количество нераспознанных растров увеличивается немонотонным образом, но в среднем на участке $[0, 16]$ увеличивается с большой скоростью, средняя скорость нарастания этой величины на участке $[16, 40]$ существенно замедляется и далее становится равной нулю, но при этом наблюдаются колебания около оценки 205,7.

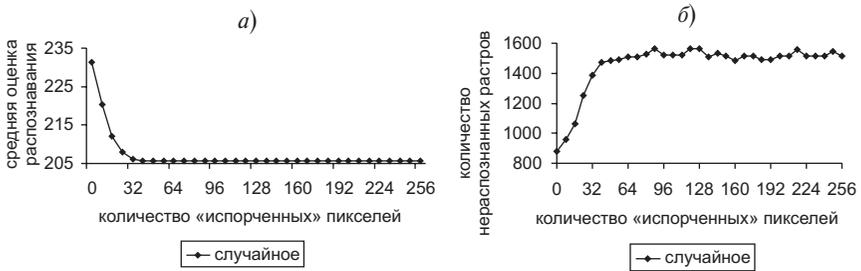


Рис. 2

Анализ поведения оценки распознавания и числа нераспознанных растров на моделях «наихудшего» и случайного разрушения позволяют нам делать выводы о поведении аналогичных характеристик для моделей разрушения, лежащих «между» двумя изученными. Очевидно, что с увеличением числа разрушенных пикселей средняя оценка распознавания будет монотонно падать, а число нераспознанных символов увеличиваться, но в среднем, немонотонно. Сделанные выводы позволяют строить следующие модели разрушения уже для всех пикселей раstra, без дополнительного анализа влияния увеличения их числа.

3. Модели затемнения (засветления)

На этапе распознавания все пиксели раstra постепенно затемняются:

$$v_i \rightarrow v_i + 0,01 \times n, \quad \text{где } n = 0, 1, 2, \dots, n_{\text{темн}}.$$

Если для каких-то пикселей начиная с некоторого n имеем: $v_i \geq 1$, то считаем, что $v_i = 1$. При засветлении аналогично

$$v_i \rightarrow v_i - 0,01 \times n, \quad \text{где } n = 0, 1, 2, \dots, n_{\text{светл}}.$$

Если для каких-то пикселей начиная с некоторого n имеем: $v_i \leq 0$, то считаем, что $v_i = 0$. При $n = n_{\text{темн}} = 12$ и аналогично при $n = n_{\text{светл}} = 20$ число неправильно распознанных растров преодолевает «пороговое» значение ($881 \times 2 = 1762$).

На рис. 3а, б для сравнения приводятся результаты распознавания, полученные при затемнении и засветлении растров. По оси абсцисс откладывается число n , определяющее степень затемнения (засветления). Распознаванию обучающей базы соответствует $n = 0$.

На рис. 3а показано почти линейное падение оценки распознавания для затемнения и засветления, причем на отрезке $[0, 12]$ средняя скорость

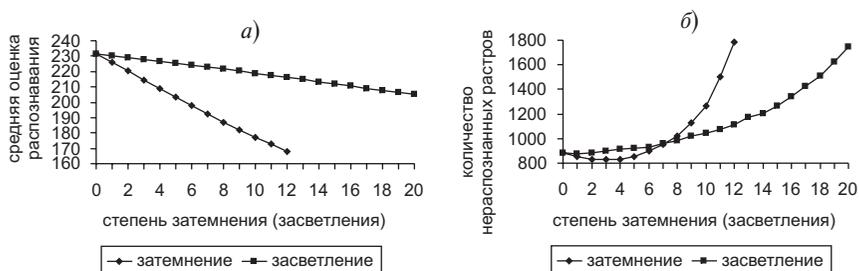


Рис. 3

второго в 4,2 раза меньше. В обоих случаях зависимости являются строго монотонными.

На рис. 3б можно видеть немонотонное изменение количества неправильно распознанных растров при затемнении. Дополнительные расчеты показали, что при $n = 2,5$ достигается минимальное число 823 неправильно распознанных символов. При $n = n_{\text{темн}} = 12$ число неправильно распознанных растров преодолевает «пороговое» значение и становится равным 1783. Для засветления график является монотонным. При $n = n_{\text{светл}} = 20$ число неправильно распознанных растров преодолевает «пороговое» значение и становится равным 1747.

Итак, затемнение и засветление растров оказывают различное влияние на распознавание. Оказывается, при небольшом затемнении (до $n = 6$) количество нераспознанных символов даже меньше, чем в немодифицированной базе. Однако, в результате наблюдающегося затем стремительного роста количества нераспознанных символов при затемнении, увеличение их числа по сравнению с распознаванием исходной базы для $n = 12$ при затемнении приблизительно в 4 раза больше, чем при засветлении. Оценка распознавания как для затемнения, так и для засветления, монотонно уменьшается с ростом степени изменения исходных растров, причем при $0 \leq n \leq 12$ для второго процесса в 4,2 раза медленнее, чем для первого.

4. Модель «дискретизации»

В рассматриваемых «серых» растрах для каждого пикселя $0 \leq v_i \leq 1$. Поделим отрезок $[0, 1]$ на 256 равных по длине частей — отрезок и 255 полуинтервалов: $[0, dv]$, $(dv, 2 \times dv]$, \dots , $(255 \times dv, 256 \times dv]$, где $dv = 1/256$. Осуществим следующее преобразование для всех пикселей растра. Если $0 \leq v_i \leq dv$, то $v_i \rightarrow dv/2$ (иначе $[0, dv] \rightarrow dv/2$); в полуинтервале $k \times dv < v_i \leq (k + 1) \times dv$, где $k = 1, \dots, 255$, производится замена: $v_i \rightarrow (k + 1/2) \times dv$ (иначе $(k \times dv, (k + 1) \times dv] \rightarrow$

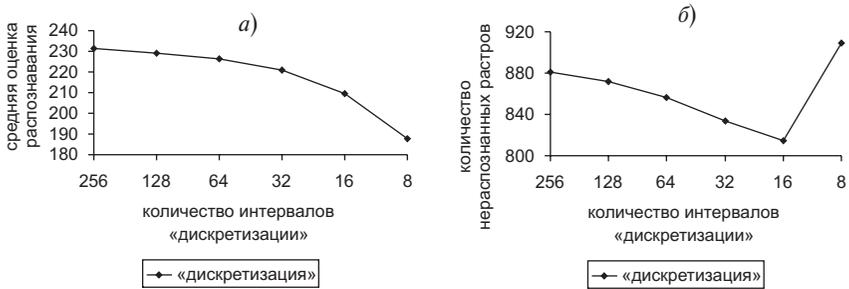


Рис. 4

$\rightarrow (k + 1/2) \times dv$). Мы осуществили дискретизацию бесконечного множества значений $0 \leq v_i \leq 1$, в результате которой v_i будет принимать только 256 значений: $\{dv/2, (1+1/2) \times dv, \dots, (255+1/2) \times dv\}$. Выполним распознавание полученной базы символов, которая, как нетрудно понять, очень незначительно отличается от исходной базы.

Произведем аналогичную дискретизацию с делением отрезка $[0, 1]$ на 128 частей, затем на 64 части, далее на 32 части, на 16, на 8 и наконец на 4. От базы к базе количество отрезков дискретизации уменьшалось в 2 раза. Каждая последующая база все больше отличается от исходной.

На рис. 4а, б приводятся результаты, полученные при «дискретизации». По оси абсцисс откладывается число отрезков дискретизации, а именно 256, 128, 64, 32, 16, 8. Распознаванию обучающей базы соответствует число 256.

На рис. 4а показано строго монотонное падение оценки распознавания при уменьшении количества отрезков дискретизации.

На рис. 4б изображен график немонотонного изменения количества неправильно распознанных растров. Дополнительные расчеты при 12 и 24 отрезках дискретизации подтвердили, что на 16 отрезках дискретизации достигается минимальное число 815 неправильно распознанных символов. Их количество увеличивается (не достигая порогового значения) при дальнейшем уменьшении числа отрезков дискретизации вплоть до 8.

На рис. 5а, б также приводятся результаты, полученные при дискретизации, но уже для числа отрезков дискретизации 256, 128, 64, 32, 16, 8, 4 (добавилось число 4).

На рис. 5а по-прежнему строго монотонное падение оценки распознавания при уменьшении количества отрезков дискретизации.

На рис. 5б изображен график немонотонного изменения количества неправильно распознанных растров. При 4 отрезках дискретизации число неправильно распознанных растров преодолевает «пороговое» значение

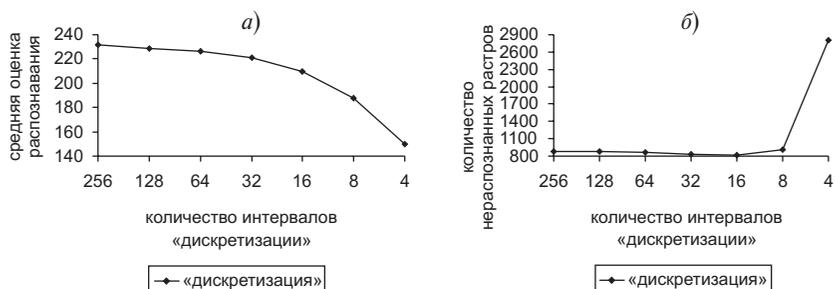


Рис. 5

и становится равным 2806. Скорость роста на отрезке [8, 4] по оси абсцисс существенно выше (в 40 раз), чем на отрезке [16, 8].

Итак, анализ результатов распознавания, полученных с использованием модели «дискретизации», показывает, что оценка распознавания монотонно уменьшается по мере «удаления» базы распознавания от базы обучения, соответствующего убыванию числа отрезков дискретизации. Количество нераспознанных символов ведет себя немонотонным образом. Вначале при уменьшении числа отрезков дискретизации оно уменьшается и достигает минимума при количестве таких отрезков, равном 16. Затем число нераспознанных символов увеличивается (причем с нарастающей скоростью) и преодолевает пороговое значение.

5. Выводы

«Превращение» базы обучения в базу распознавания формализовано в виде четырех моделей. Для каждой из них получено, что по мере увеличения степени «искажения» обучающей базы оценка распознавания монотонно уменьшается, а количество нераспознанных символов в целом увеличивается, но этот процесс может протекать существенно немонотонным образом.

При реальном распознавании можно использовать полученные наработки для улучшения качества распознавания, но уже в «обратном» порядке. Разумеется, имеются в виду только те модели, что применяются для всех пикселей раstra, а не для некоторых случайным образом выбранных. Предложенные способы преобразования (затемнение, засветление, «дискретизацию») можно применить к распознаваемой последовательности символов. Из вышеописанного следует, что результатом может быть либо улучшение оценок распознавания, либо уменьшение числа нерас-

познанных символов, либо и то, и другое. Возможно комбинированное использование затемнения (засветления) и «дискретизации».

Литература

1. *Гавриков М. Б., Пестрякова Н. В.* Метод полиномиальной регрессии в задачах распознавания печатных и рукопечатных символов: Препринт // ИПМатем. РАН. М., 2004. № 22. 12 с.
2. *Гавриков М. Б., Пестрякова Н. В., Славин О. А., Фарсобина В. В.* Развитие метода полиномиальной регрессии и практическое применение в задаче распознавания: Препринт // ИПМатем. РАН. М., 2006. № 25. 21 с.
3. *Гавриков М. Б., Мисюрев А. В., Пестрякова Н. В., Славин О. А.* Об одном методе распознавания символов, основанном на полиномиальной регрессии // Автоматика и Телемеханика. 2006. № 2. С. 119–134.
4. *Гавриков М. Б., Пестрякова Н. В., Усков А. В., Фарсобина В. В.* О некоторых свойствах метода распознавания символов, основанного на полиномиальной регрессии: Препринт // ИПМатем. РАН. М., 2007. № 69. 20 с.
5. *Гавриков М. Б., Пестрякова Н. В., Усков А. В., Фарсобина В. В.* О некоторых свойствах оценки метода распознавания символов, основанного на полиномиальной регрессии: Препринт // ИПМатем. РАН. М., 2008. № 7. 28 с.
6. *Гавриков М. Б., Пестрякова Н. В., Усков А. В., Фарсобина В. В.* Оценка распознавания символов для метода, основанного на полиномиальной регрессии // Настоящий сборник. С. 194–219.