

Зависимость качества распознавания от степени различия между базами обучения и распознавания

Н. В. Пестрякова

*Институт системного анализа Российской академии наук,
Россия, 117312 Москва, пр. 60-летия Октября, 9*

Для метода распознавания символов, основанного на регрессионном анализе, разработан подход к исследованию зависимости качества распознавания от степени различия между базой обучения и полученной путем ее модификации базой распознавания. В процессе изменения базы рукопечатных цифр изучена корреляция динамики средне-статистических величин (растров и полиномиальных векторов), а также качества распознавания.

Введение

Исследуется зависимость качества распознавания от степени различия между базами обучения (рис. 1) и распознавания. Интуитивно понятно, что любая база распознавания представляет собой каким-либо образом «испорченную» базу обучения. На практике бывает трудно определить, существует ли некая закономерность этого «превращения». Было предложено использовать в качестве базы распознавания заданным образом модифицированную базу обучения. Анализируются три модели такого преобразования, а именно: «затемнения», «засветления» и «дискретизации».

При нарастании степени искажения исходной базы рукопечатных цифр изучается динамика оценки распознавания (средней по всем верно распознанным изображениям), а также числа неправильно распознанных образов. Ставится задача выявить особенности корреляции в поведении указанных величин и базирующихся на полиномиальных векторах и растрах [1–4] характеристик базы распознавания.

1. Модели «затемнения» («засветления»)

1.1. Описание моделей и некоторые замечания

На этапе распознавания все пиксели растра постепенно затемняются: $v_i \rightarrow v_i + 0,01 \cdot n$, где $n = 0, 1, \dots, 100$. Если для каких-то пикселей начиная с некоторого n имеем: $v_i > 1$, то считаем $v_i = 1$. При засветлении аналогично $v_i \rightarrow v_i - 0,01 \cdot n$, где $n = 0, 1, \dots, 100$. Если для каких-то пикселей начиная с некоторого n имеем $v_i < 0$, то считаем $v_i = 0$.

В табл. 1 приведены результаты расчетов не для всех целых неотрицательных n , принадлежащих исследуемому отрезку $0 \leq n \leq 100$. Выбор n определялся условием, чтобы табличные данные достаточно подробно отражали выявленные закономерности (поэтому шаг по n непостоянен). Графики на рис. 2–9 построены с использованием только части результатов из табл. 1, а именно, для значений n , кратных 8. В тексте при ссылке на графики специально оговариваются случаи, когда по сравнению с табличными данными имеется потеря информации (незначительная).

В табл. 1 утолщением и затемнением выделены экстремальные значения. Понятно, что табличные данные являются округленными. Реальные экстремумы могут находиться на шаге левее или правее указанного n и несколько отличаться от табличного значения.

1.2. Качество распознавания и интегральные характеристики модифицированной базы. Сравнительный анализ при затемнении и засветлении

В табл. 1 для изучаемого диапазона $0 \leq n \leq 100$ содержатся данные для величин, осредненных по совокупности изображений символов. Представлено отклонение между среднестатистическими растрами баз обучения и распознавания (vv_t при затемнении и vv_s при засветлении) — см. также рис. 7, кроме того, приведено отклонение между среднестатистическими полиномиальными векторами баз обучения и распознавания (xx_t при затемнении и xx_s при засветлении) — см. также рис. 4. Все четыре величины получены независимо от правильности распознавания, поскольку их следует рассматривать как некие интегральные характеристики для базы распознавания в целом. Именно таким образом (но для каждого из символов в отдельности) среднестатистические растры/векторы были введены в работах [1, 2]. Там же было показано, что эти величины являются ключевыми характеристиками базы.

Очевидно, что для произвольного символа как при затемнении, так и при засветлении рост n сопровождается увеличением расстояния между среднестатистическими растрами в базах обучения и распознавания (среднее по совокупности символов — рис. 7). Для среднестатистических полиномиальных векторов при затемнении сохраняется указанная законо-

мерность, а при засветлении имеется небольшой диапазон степеней n , где это расстояние незначительно уменьшается (среднее по совокупности символов — рис. 4).

В табл. 1 для $0 \leq n \leq 100$ приведены число нераспознанных символов (mis_t при затемнении и mis_s при засветлении) — рис. 3, а также средняя оценка распознавания (Prb_t при затемнении и Prb_s при засветлении) — рис. 2.

Сравним поведение указанных величин при затемнении и засветлении.

Во всем исследуемом диапазоне ($0 \leq n \leq 100$) функции vv_t и vv_s гладкие и почти линейные, с небольшим замедлением темпов их роста при увеличении n ; кроме того, для каждого n величина vv_t более чем в два раза превышает vv_s (рис. 7) — объяснение приведенных фактов дано в п. 1.11. В соответствии с этим, для каждого n величина xx_t более чем в полтора раза превышает xx_s (рис. 4). В согласии с отмеченными закономерностями, для количества неправильно распознанных символов при любом n из указанного интервала имеем: значение mis_t в полтора–два раза больше, чем mis_s (рис. 3).

Средняя оценка распознавания и при затемнении, и при засветлении сначала уменьшается, а затем увеличивается (рис. 2). Но для Prb_t средние темпы и падения, и роста в полтора–два раза выше, чем для Prb_s (см. п. 1.5). При затемнении минимум средней оценки распознавания достигается при меньшем n , чем при засветлении; кроме того, минимальное значение Prb_s меньше, чем минимальное значение Prb_t , однако их различие незначительно.

1.3. Корреляция качества распознавания и интегральной характеристики модифицированной базы при засветлении

На рис. 4 видно, что график xx_s «перестраивается» при $64 \leq n \leq 72$. На самом этом отрезке xx_s убывает, правда, очень незначительно, а левее и правее указанного участка xx_s нарастает с почти постоянной скоростью, причем для $n > 72$ она приблизительно в два раза ниже, чем при $n < 64$.

На этом же отрезке $64 \leq n \leq 72$ «перестраивается» и график, соответствующий изменению количества неправильно распознанных символов mis_s . На указанном участке mis_s убывает (незначительно, по сравнению с количеством неправильно распознанных символов). Слева и справа от этого отрезка mis_s монотонно увеличивается. Темп роста mis_s повышается по мере приближения к указанному участку слева. Справа от него mis_s увеличивается с почти постоянной скоростью, которая ниже, чем соответствующая величина в непосредственной близости от отрезка $64 \leq n \leq 72$ слева.

Исключая из рассмотрения локальное и очень небольшое нарушение монотонности величины mis_s на начальном участке (минимум при $n = 1$) —

Таблица 1

Засветление и затемнение

	0	1	2	3	4	8	12	16	20	24	32	40	48	56	64	72	80	88	96	100
dv_tw	60,7	59,6	58,8	58,2	57,8	57,2	57,9	59,7	62,6	66,2	75,7	87,3	100,3	113,6	126,8	140,2	154,5	170,4	183,5	189,9
dv_sw	60,7	60,2	59,8	59,3	58,9	57,3	55,8	54,5	53,4	52,5	51,6	51,7	52,7	54,2	56,4	59,7	62,8	65,8	68,6	69,7
dv_t	60,7	59,7	58,9	58,3	57,9	57,2	58,0	59,9	62,7	66,4	75,8	87,4	100,3	113,7	127,3	140,8	153,8	166,3	178,1	183,8
dv_s	60,7	60,3	59,8	59,4	59,0	57,4	55,9	54,7	53,7	52,8	52,0	52,2	53,6	55,8	58,5	61,7	65,1	68,3	71,1	72,2
dv_db	76,4	75,1	74,0	73,2	72,3	68,8	66,7	66,2	67,4	70,3	76,3	87,7	100,3	113,8	127,6	141,0	153,5	165,3	176,9	182,3
dv_sb	76,4	75,7	75,2	74,7	74,1	72,3	72,2	72,9	70,3	67,3	63,4	61,9	62,0	62,9	64,4	67,3	70,0	72,0	74,0	74,8
vv_t	0	2,3	4,6	6,8	9,1	18,0	27,0	35,2	43,5	51,6	67,9	82,7	98,6	112,7	127,2	140,7	153,8	166,3	178,1	183,8
vv_s	0	1,2	2,3	3,5	4,6	9,0	13,2	17,3	21,2	24,9	32,1	38,7	45,0	50,0	55,8	60,5	64,6	68,2	71,1	72,2
dx_tw	3220	3188	3166	3151	3141	3140	3175	3231	3297	3365	3697	3872	4072	4307	4570	4875	5262	5752	6285	6585
dx_sw	3220	3205	3191	3176	3165	3115	3072	3036	3012	2998	3009	3061	3147	3253	3381	3452	3549	3655	3764	3814
dx_t	3224	3192	3170	3155	3145	3144	3180	3238	3304	3373	3720	3904	4115	4362	4651	4988	5375	5818	6322	6598
dx_s	3224	3209	3195	3180	3169	3120	3078	3046	3026	3014	3030	3094	3202	3343	3503	3573	3689	3808	3922	3974
dx_db	4064	4020	3980	3956	3919	3783	3671	3593	3560	3575	3819	3989	4181	4412	4698	5035	5412	5833	6331	6600
dx_sb	4064	4047	4037	4027	4012	3966	3986	4052	3970	3875	3757	3713	3722	3764	3841	3915	3974	4040	4107	4142
xx_t	0	76	150	222	292	558	805	1036	1252	1458	2568	2860	3154	3460	3790	4153	4558	5013	5523	5800
xx_s	0	44	86	128	175	344	513	680	850	1013	1352	1693	2038	2384	2705	2703	2857	2999	3123	3176
mis_t	881	882	891	905	927	1200	1843	2952	4772	6972	33309	49051	69225	91758	110884	123073	131767	141501	141617	141617
mis_s	881	871	884	892	907	1019	1279	1848	2523	3223	5077	8835	16860	30896	46397	45692	57590	69416	80611	85241
Ptb_t	231,3	229,1	226,6	224,0	221,5	211,5	202,2	193,9	186,4	179,3	172,5	174,0	182,5	200,5	224,4	244,1	250,5	255,0	255,0	255,0
Ptb_s	231,3	230,6	229,8	228,9	228,1	224,1	219,7	215,2	210,3	205,4	194,7	184,2	175,3	170,6	170,1	174,9	178,3	182,5	187,1	189,0

см. табл. 1, в целом следует отметить наличие следующей закономерности: при $n < 64$ и $n > 72$ нарастанию xx_s соответствует увеличение mis_s , а при $64 \leq n \leq 72$ очень незначительно убывают xx_s , и mis_s .

Именно в области $64 \leq n \leq 72$ поведение средней оценки распознавания Ptb_s коренным образом меняется от убывания к нарастанию: при $n = 64$ наблюдается минимум этой величины.

1.4. Корреляция качества распознавания и интегральной характеристики модифицированной базы при затемнении

На рис. 4 видно, что график xx_t является монотонно возрастающим и кусочно-линейным. Точки, где меняются углы наклона (рвутся производные), соответствуют $n = 24$ и $n = 32$. На отрезке $24 \leq n \leq 32$ скорость нарастания величины xx_t приблизительно в три раза больше, чем левее и правее указанного участка (где они примерно одинаковы).

На этом же отрезке $24 \leq n \leq 32$ радикально «перестраивается» график, соответствующий изменению количества неправильно распознанных символов mis_t . На указанном участке скорость увеличения mis_t выше, чем на любой части отрезка $0 \leq n \leq 24$ (и приблизительно в десять раз больше, чем в среднем на этом отрезке). При $32 \leq n \leq 100$ значения mis_t монотонно увеличиваются, но темпы роста на любой части этого отрезка меньше, чем при $24 \leq n \leq 32$. Кроме того, следует отметить, что вблизи $n = 0$ и $n = 100$ скорости увеличения mis_t низкие.

В общем, наблюдается следующая закономерность: в диапазоне $0 \leq n \leq 100$ монотонно возрастают и xx_t , и mis_t . При $24 \leq n \leq 32$ темпы увеличения указанных величин намного выше, чем на остальных участках.

Именно в области $24 \leq n \leq 32$ поведение средней оценки распознавания Ptb_t коренным образом меняется от убывания к нарастанию: при $n = 32$ имеется минимум этой величины.

1.5. О сильном затемнении и засветлении

Конечно же, при больших значениях n нельзя говорить о сколько-нибудь приемлемом распознавании, и результаты носят чисто теоретический характер. Следует отметить, что ситуация в этой области асимметрична для затемнения и засветления. А именно, при полном затемнении с $n = 100$ («Черный квадрат») все символы распознаются как «1» с максимальной оценкой 255. Этим и объясняется стабильное нарастание средней оценки распознавания Ptb_t для $n > 32$ при затемнении. Однако при полном засветлении с $n = 100$ ситуация не такая замечательная, поскольку среди цифр от 0 до 9 нет столь уникально подходящей для «Белого квадрата». Соответственно наблюдается нестабильность: в качестве искомого решения выбираются различные цифры с существенно более низкими оценками, чем при

Таблица 2

Распознавание «Белого квадрата»

	0	1	2	3	4	5	6	7	8	9
0	16645	17	9	1184	2	1721	0	1	0	0
1	0	29428	460	3002	1	176	93	1	0	0
2	0	60	17576	233	0	121	0	0	0	0
3	0	71	491	12462	0	574	44	10	0	0
4	8	2100	3715	5408	1177	455	31	8	0	0
5	0	5	17	4786	0	8310	8	1	0	0
6	3	157	315	10583	0	1869	119	0	0	0
7	2	376	1346	7047	32	3139	400	3809	0	0
8	0	162	2529	6837	0	577	17	1	0	0
9	12	68	4358	19285	0	1277	9	27	0	11
Σ	16670	32444	30816	70827	1212	18219	721	3858	0	11

полном затемнении. Иллюстрацией к сказанному является табл. 2, в которой при $n = 100$ предлагаемым для распознавания одинаковым изображениям («Белый квадрат») разных символов, а также их совокупности (перечисленные в левом столбце цифры 0, 1, ..., 9 и знак Σ) приводится в соответствии количество случаев, в которых данное изображение этого символа или их совокупности распознается как один из символов верхней строки 0, 1, ..., 9.

1.6. Дифференциальные характеристики модифицированной базы

В табл. 1 приведены среднее по совокупности символов отклонение растров базы распознавания от среднестатистического растра базы обучения (см. также рис. 8, 9) и, аналогично, среднее по совокупности символов отклонение полиномиальных векторов базы распознавания от среднестатистического вектора базы обучения (см. также рис. 5, 6). Они были введены в работе [3] для каждого из символов в отдельности. В данной работе для каждого изображения некоторого символа определяется отклонение от среднестатистического растра или вектора этого символа, а затем находится среднее значение по всем изображениям рассматриваемой совокупности символов по базе.

Эти величины в табл. 1 даны без учета правильности распознавания (см. также рис. 5, 8), для правильного распознавания (аналогично изображенному на рис. 5, 8) и, наконец, для неправильно распознанных символов (см. также рис. 6, 9). Используемые обозначения представлены в табличном виде (табл. 3).

Таблица 3

Дискретизация

	Затемнение	Засветление
растры без учета правильности распознавания	dv_t	dv_s
растры, правильное распознавание	dv_tw	dv_sw
растры, неправильное распознавание	dv_tb	dv_sb
векторы без учета правильности распознавания	dx_t	dx_s
векторы, правильное распознавание	dx_tw	dx_sw
векторы, неправильное распознавание	dx_tb	dx_sb

Казалось бы, все эти отклонения от среднестатистических растров/векторов с ростом n должны увеличиваться (аналогично vv_t , vv_s , xx_t) или с некоторыми погрешностями (как xx_s). Однако расчеты показали нечто иное.

1.7. Динамика дифференциальных характеристик модифицированной базы при засветлении

При усилении засветления dv_s и dv_sw уменьшаются и достигают минимума при $n = 32$, после чего увеличиваются. Для всего рассматриваемого отрезка изменения n ($0 \leq n \leq 100$) различия dv_s и dv_sw незначительны, поэтому приведен график только для первой из них (рис. 8). С ростом n величина dv_sb сначала уменьшается (за исключением участка незначительного нарастания при $12 \leq n \leq 16$) и достигает минимума при $n = 40$, а затем увеличивается (рис. 9). Следует заметить, что при $0 \leq n \leq 100$ величина dv_sb существенно больше, чем dv_s и dv_sw ; и кроме того, как нетрудно понять, из неравенства $dv_sb > dv_sw$ следует неравенство $dv_s > dv_sw$.

При анализе поведения аналогичных величин в терминах полиномиальных векторов оказывается, что при усилении засветления dx_s и dx_sw уменьшаются и достигают минимума при $n = 24$, после чего увеличиваются. Для всего рассматриваемого интервала изменения n ($0 \leq n \leq 100$) различия dx_s и dx_sw незначительны, поэтому приведен график только для первой из них (рис. 5). Величина dx_sb уменьшается (за исключением участка незначительного нарастания при $8 \leq n \leq 16$) и достигает минимума при $n = 40$, после чего увеличивается, оставаясь существенно больше, чем dx_s и dx_sw на всем исследуемом диапазоне изменения n ($0 \leq n \leq 100$) (рис. 6). Кроме того, из неравенства $dx_sb > dx_sw$ следует неравенство $dx_s > dx_sw$.

Итак, при усилении засветления поведение величин dv_sb , dv_sw и dv_s в значительной степени сходно с поведением dx_sb , dx_sw и dx_s .

В качестве критерия того, что одна величина существенно больше другой, а какие-то величины сопоставимы, для определенности использовалось сравнение с 5 % от их среднего арифметического.

1.8. Корреляция качества распознавания, интегральной и дифференциальных характеристик модифицированной базы при засветлении

Следует отметить такую закономерность. На отрезке $64 \leq n \leq 72$ радикально «перестраиваются» графики отклонения между среднестатистическими полиномиальными векторами баз обучения и распознавания xx_s , количества неправильно распознанных символов mis_s , а также средней оценки распознавания Prb_s (п. 1.3). Именно вблизи этого участка после этапа убывания возрастают до своих начальных значений $dv_sb, dv_sw, dv_s, dx_sw, dx_s$. Более точно, с dx_sw, dx_s это происходит с немного меньшими значениями n , с dv_sw, dv_s на самом этом участке, а dv_sb с немного большими значениями n . Величина dx_sb приближается к исходному значению, но не достигает его даже при $n = 100$.

1.9. Динамика дифференциальных характеристик модифицированной базы при затемнении

При усилении затемнения расстояния dv_t и dv_tw сначала уменьшаются и достигают минимума при существенно меньшем значении n по сравнению с засветлением ($n = 8$), а затем возрастают. Величины dv_t и dv_tw отличаются очень мало на всем исследуемом диапазоне значений n ($0 \leq n \leq 100$), поэтому приведен график только для первой из них (рис. 8). С ростом n величина dv_tb сначала уменьшается и достигает минимума при существенно меньшем значении n по сравнению с засветлением ($n = 16$), а затем увеличивается (рис. 9). Следует заметить, что при $0 \leq n \leq 16$ величина dv_tb существенно больше, чем dv_t и dv_tw ; кроме того, из неравенства $dv_tb > dv_tw$ следует неравенство $dv_t > dv_tw$. При $n = 20$ dv_tb становится сопоставимым с dv_tw и dv_t , и дальнейшее нарастание dv_tb, dv_t и dv_tw идет уже «синхронно», причем если $dv_tb < dv_tw$, то $dv_t < dv_tw$ (при $n \geq 80$), если $dv_tb = dv_tw$, то $dv_t = dv_tw$ (при $n = 48$) а если $dv_tb > dv_tw$, то $dv_t > dv_tw$ (при $n < 80, n \neq 48$).

Если проанализировать поведение аналогичных величин в терминах полиномиальных векторов, то оказывается, что при усилении затемнения расстояния dx_t и dx_tw сначала уменьшаются и достигают минимума при $n = 8$, а затем увеличиваются. Величины dx_t и dx_tw отличаются очень мало на всем исследуемом диапазоне значений n ($0 \leq n \leq 100$), поэтому приведен график только для первой из них (рис. 5). С ростом n величина dx_tb сначала уменьшается и достигает минимума при $n = 20$, а затем уве-

личивается (рис. 6). Следует заметить, что при $0 \leq n \leq 96$ величина dx_{tb} существенно больше, чем dx_t и dx_{tw} ; кроме того, из неравенства $dx_{tb} > dx_{tw}$ следует неравенство $dx_t > dx_{tw}$. При $n = 100$ значение dx_{tb} становится сопоставимым с dx_{tw} и dx_t , причем $dx_{tb} > dx_{tw}$, и следовательно, $dx_t > dx_{tw}$.

Следует еще раз подчеркнуть тот факт, что для базы обучения ($n = 0$) среднее отклонение растров неправильно распознанных символов от среднестатистического растра (dv_{tb}) существенно больше, чем среднее отклонение растров правильно распознанных символов от среднестатистического растра (dv_{tw}), но при $20 \leq n \leq 100$ они уже различаются мало. В то же время для всех n (кроме $n = 100$) среднее отклонение векторов неправильно распознанных символов от среднестатистического полиномиального вектора (dx_{tb}) существенно больше, чем среднее отклонение векторов правильно распознанных символов от среднестатистического вектора (dx_{tw}). Следовательно, при затемнении для векторов имеет место структурированное расположение символов (отсутствующее для растров).

В остальном при усилении затемнения поведение величин dv_{tb} , dv_{tw} , dv_t в значительной степени сходно с поведением dx_{tb} , dx_{tw} , dx_t .

1.10. Корреляция качества распознавания, интегральной и дифференциальных характеристик модифицированной базы при затемнении

Следует отметить еще одну важную закономерность. На отрезке $24 \leq n \leq 32$ радикально «перестраиваются» графики отклонения между среднестатистическими полиномиальными векторами баз обучения и распознавания xx_t , количества неправильно распознанных символов mis_t , а также средней оценки распознавания Prb_t (п. 1.4). Именно вблизи этого участка после убывания возрастают до своих начальных значений dv_{tb} , dv_{tw} , dv_t , dx_{tb} , dx_{tw} , dx_t ; точнее, с dv_{tw} , dv_t , dx_{tw} , dx_t это происходит с немного меньшими значениями n , с dv_{tb} на самом этом участке, а dx_{tb} с немного большими значениями n .

1.11. Расстояние между растрами «Белого» и «Черного» квадратов

По табл. 1 видно, что при $n = 100$ выполняются равенства: $vv_t = dv_t = 183,8$ (отклонение растра «Черного квадрата» от среднестатистического растра обучающей базы, причем первое равенство легко получить аналитически), а также $vv_s = dv_s = 72,2$ (отклонение растра «Белого квадрата» от среднестатистического растра обучающей базы, первое равенство получается аналитически). Следовательно, $vv_t + vv_s = dv_t + dv_s = 256$ (все эти равенства легко получить аналитически). Указанная сумма соответствует расстоянию между растрами «Белого» и «Черного» квадратов.

Итак, отклонение среднестатистического растра базы обучения от растра «Черного» квадрата в два с половиной раза больше, чем от растра «Белого» квадрата. Этим объясняется аналогичное различие в темпах роста vv_t и vv_s .

Скорость роста обеих указанных функций замедляется с ростом n , поскольку при этом увеличивается количество пикселей, в которых достигается значение «1» при затемнении и «0» при засветлении.

2. Модель «дискретизации»

2.1. Описание модели. Представление дискретизации в виде комбинации затемнения и засветления

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: $[0, dv]$, $(dv, 2 \cdot dv]$, ..., $(255 \cdot dv, 256 \cdot dv]$, где $dv = 1/256$. Осуществим для всех пикселей растра следующее преобразование: если $0 \leq v_i \leq dv$, то $v_i \rightarrow dv/2$ (иначе $[0, dv] \rightarrow dv/2$); в полуинтервале $k \cdot dv < v_i \leq (k+1) \cdot dv$, где $k = 1, \dots, 255$, производится замена: $v_i \rightarrow (k+1/2) \cdot dv$ (иначе $(k \cdot dv, (k+1) \cdot dv] \rightarrow (k+1/2) \cdot dv$). Тем самым реализуется дискретизация бесконечного множества значений $0 \leq v_i \leq 1$, в результате которой v_i может принимать только 256 значений: $\{dv/2, (1+1/2) \cdot dv, \dots, (255+1/2) \cdot dv\}$. Выполним распознавание полученной базы символов, которая очень незначительно отличается от исходной базы.

Произведем аналогичную дискретизацию с делением отрезка $[0, 1]$ на различное количество отрезков дискретизации $N_{\text{дискр}}$, а именно, на 128 частей, затем на 64 части, далее на 32, 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.

Нетрудно понять, что дискретизация является комбинацией затемнения и засветления. Для заданного количества отрезков дискретизации $N_{\text{дискр}}$ получим, какое максимальное n может быть в рамках моделей «затемнения» и «засветления». В данном анализе будем считать n , как и раньше, неотрицательным, но не целым, как в п. 1, а действительным. От этого не возникает никаких принципиальных расхождений с изложенным в п. 1, поскольку там тоже можно было считать n действительным, и только для простоты использовались целые значения.

В рамках модели «дискретизации» в каждом пикселе растра выполняется следующее преобразование (для диапазона, внутри которого находится значение в пикселе): либо $(k \cdot dv, (k+1) \cdot dv] \rightarrow (k+1/2) \cdot dv$ для $k = 1, \dots, N_{\text{дискр}} - 1$, либо $[0, dv] \rightarrow dv/2$ для $k = 0$, причем $dv = 1/N_{\text{дискр}}$. Это можно представить иначе (для значения в пикселе): $F \cdot dv \rightarrow (k+1/2) \cdot dv$, причем либо $k < F \leq (k+1)$ для $k = 1, \dots, N_{\text{дискр}} - 1$, либо $0 \leq F \leq 1$ для $k = 0$. В рамках

модели «затемнения» используется следующее превращение (для значений в пикселе): $v_i \rightarrow v_i + 0,01 \cdot n$ (для «засветления» $v_i \rightarrow v_i - 0,01 \cdot n$), где $n = 0, 1, 2, \dots$. Рассматриваются только такие пиксели, в которых после затемнения $v_i \leq 1$ и после засветления $v_i \geq 0$. Абсолютная величина разницы модифицированного и исходного значений в пикселях для затемнения и засветления при одном и том же n равны $0,01 \cdot n$. Для дискретизации эту величину можно представить так: $|(k + 1/2) - F| \cdot dv$, причем либо $k < F \leq (k + 1)$, где $k = 1, \dots, N_{\text{дискр}} - 1$, либо $0 \leq F \leq 1$ для $k = 0$. Итак, $0,01 \cdot n = |(k + 1/2) - F| \cdot dv = |(k + 1/2) - F| / N_{\text{дискр}}$. Получим ограничение для величины в правой части равенства: $|(k + 1/2) - F| / N_{\text{дискр}} \leq 1/(2 \cdot N_{\text{дискр}})$. Следовательно, $0,01 \cdot n \leq 1/(2 \cdot N_{\text{дискр}})$. Окончательный результат: $n \leq 50/N_{\text{дискр}}$. В табл. 5 представлены значения величины $f_{\text{дискр}} = 50 / N_{\text{дискр}}$ в зависимости от $N_{\text{дискр}}$.

2.2. Анализ табличных и графических данных. Согласованность характеристик качества распознавания для дискретизации с соответствующими величинами для затемнения и засветления

В табл. 5 содержатся данные для осредненных величин по совокупности символов при $N_{\text{дискр}} = 256, 128, 64, 32, 16, 8, 4$. Приведены отклонения между среднестатистическими растрами vv и между среднестатистическими полиномиальными векторами xx баз обучения и распознавания. Обе величины получены независимо от правильности распознавания, как и в п. 1. В табл. 5 также приводятся число нераспознанных символов mis и средняя оценка распознавания Prb .

Очевидно, что для произвольного символа уменьшение $N_{\text{дискр}}$ должно сопровождаться увеличением расстояния между среднестатистическими растрами в базах обучения и распознавания vv (рис. 15). Для среднестатистических полиномиальных векторов xx сохраняется указанная закономерность (рис. 12). Заметим, что темпы роста vv и xx , незначительные вблизи $N_{\text{дискр}} = 256$, нарастают с уменьшением $N_{\text{дискр}}$ — графики являются вогнутыми.

Особенность рис. 12–17 в том, что расстояния на оси абсцисс не соответствуют меткам делений — цифрам 256, 128, 64, 32, 16, 8, 4, которые образуют убывающую геометрическую прогрессию с коэффициентом 1/2. Для того чтобы установить это соответствие, следует длину отрезка [128, 64] сделать в два раза меньше, чем для отрезка [256, 128], длину отрезка [64, 32] сделать в два раза меньше, чем для отрезка [128, 64], и т. д. Это аналогично сжатию вдоль этой оси фрагментов графиков для каждого последующего отрезка по сравнению с предыдущим в два раза. При этом графики станут еще более крутыми.

Можно заменить метки делений $N_{\text{дискр}}$ на их «предельные» степени затемнения/засветления, а именно, $f_{\text{дискр}} = 50/N_{\text{дискр}}$. Расстояния на оси абсцисс не будут соответствовать новым меткам делений — числам

50/256, 50/128, 50/64, 50/32, 50/16, 50/4, которые образуют возрастающую геометрическую прогрессию с коэффициентом 2. Чтобы установить это соответствие, следует длину отрезка [50/128, 50/64] сделать в два раза больше, чем для отрезка [50/256, 50/128], длину отрезка [50/64, 50/32] сделать в два раза больше, чем для отрезка [50/128, 50/64], и т. д. Это соответствует растяжению вдоль этой оси фрагментов графиков для каждого отрезка по сравнению с предыдущим в два раза. Графики станут более полными и сходными с полученными при затемнении/засветлении.

Дальнейший анализ проводится для шкалы, изображенной на рис. 12–17. Для линейной по $N_{\text{дискр}}$ шкалы он качественно не изменится, а скорости или темпы роста, о которых будет говориться далее, увеличатся. Сравнение с затемнением и засветлением будет проведено отдельно и более точно, с использованием табличных данных.

Рост (с ускорением) vv (рис. 15) и xx (рис. 12) при уменьшении $N_{\text{дискр}}$ соответствует уменьшению средней оценки распознавания Prb (рис. 10), причем темпы убывания последней также увеличиваются с уменьшением $N_{\text{дискр}}$. При этом количество неправильно распознанных символов mis (рис. 11) с уменьшением числа отрезков дискретизации от 256 до 128 очень незначительно увеличивается, от 128 до 64 — уменьшается, а при дальнейшем падении числа отрезков дискретизации от 64 до 4 — увеличивается с нарастающими темпами.

Отметим, что указанные в предыдущем абзаце графики качественно соответствуют аналогичным для затемнения и засветления с диапазоном степеней n $0 \leq n \leq 50/4$.

Величина mis достигает минимального значения при $N_{\text{дискр}} = 64$, для которого $f_{\text{дискр}} = 0,8$. Ограничение на степени затемнения/засветления будет следующим: $n \leq 0,8$. По табл. 1 видно, что при $n = 1$ для засветления также зафиксирован близкий по значению минимум mis_s . Нетрудно проверить, что для любого $N_{\text{дискр}}$ из всего исследуемого набора интервалов дискретизации выполняется следующее: отклонение между среднестатистическими растрами / полиномиальными векторами баз обучения и распознавания, количество нераспознанных символов и средняя оценка распознавания лежат в диапазоне между соответствующими результатами для затемнения и засветления при степени $n = f_{\text{дискр}}$ (или близки к этим результатам).

В табл. 5 приведены среднее отклонение растров базы распознавания от среднестатистического растра базы обучения (рис. 16, 17), а также среднее отклонение полиномиальных векторов базы распознавания от среднестатистического вектора базы обучения (рис. 13, 14). Эти величины в табл. 5 даны без учета правильности распознавания (рис. 13, 16), для правильно распознанных (аналогично рис. 13, 16) и для неправильно распознанных символов (рис. 14, 17). Используемые обозначения представлены в табличном виде (табл. 4).

Таблица 4

растры без учета правильности распознавания	dv
растры, правильное распознавание	dvw
растры, неправильное распознавание	dvb
векторы без учета правильности распознавания	dx
векторы, правильное распознавание	dxw
векторы, неправильное распознавание	dxb

Таблица 5

	256	128	64	32	16	8	4
dvw	60,3	60,0	59,4	58,4	56,7	54,1	51,0
dv	60,4	60,1	59,5	58,5	56,7	54,1	51,1
dvb	75,9	75,7	75,1	73,3	70,6	66,3	59,1
dxw	3210	3199	3181	3156	3106	3022	2965
dx	3214	3203	3185	3160	3110	3027	2972
dxb	4051	4041	4025	3972	3887	3725	3438
vv	0,25	0,5	1,0	2,0	4,0	8,0	16,3
xx	10,1	22,2	41,3	82,2	155,2	310,7	563,5
mis	875	879	868	881	915	1081	2906
Prb	230,7	230,2	229,0	226,6	221,1	209,3	189,3
$f_{\text{дискр}}$	0,2	0,4	0,8	1,6	3,1	6,3	12,5

Расчеты показали, что при уменьшении $N_{\text{дискр}}$ расстояния dvb, dvw, dv, dxw, dxb, dx монотонно уменьшаются с увеличивающимися темпами. На всем исследуемом диапазоне значений $N_{\text{дискр}}$ ($4 \leq N_{\text{дискр}} \leq 256$) имеют место следующие закономерности: величины dv и dvw отличаются очень мало; величина dvb существенно больше, чем dv и dvw (аналогично затемнению и засветлению для $0 \leq n \leq 12,5$); из неравенства $dvb > dvw$ следует неравенство $dv > dvw$. В терминах полиномиальных векторов величины dx и dxw отличаются очень мало; величина dxb существенно больше, чем dx и dxw (аналогично затемнению и засветлению для $0 \leq n \leq 12,5$); из неравенства $dxb > dxw$ следует неравенство $dx > dxw$.

Выводы

«Преобразование» базы обучения в базу распознавания формализовано в виде трех моделей, а именно: «затемнения», «засветления» и «дискретизации».

При увеличении степени затемнения/засветления расстояние между среднестатистическими растрами баз обучения и распознавания увеличивается почти линейно, с небольшим падением темпов роста.

Расстояние между среднестатистическими векторами изменяется кусочно-линейно, график состоит из трех частей, причем в крайних частях наблюдается монотонное возрастание, а в центральной при затемнении — стремительный рост, при засветлении — незначительное убывание.

Количество неправильно распознанных изображений изменяется аналогично расстоянию между среднестатистическими векторами (будучи при этом нелинейной функцией); исключением является небольшое уменьшение числа нераспознанных символов на начальном этапе засветления.

Для любой степени затемнения каждая из этих величин больше, чем для такого же уровня засветления.

Средняя оценка распознавания и при затемнении, и при засветлении сначала убывает, а затем возрастает (при затемнении скорости больше). Минимумы приблизительно равны по величине и приходятся на центральные части двух указанных ранее зависимостей.

Посчитанные для базы обучения средние отклонения растров/векторов от среднестатистических растров/векторов «очерчивают» вокруг последних некоторую окрестность. Для засветления, как и для затемнения при переходе через эту окрестность «перестраиваются» графики отклонения между среднестатистическими полиномиальными векторами баз обучения и распознавания, количества неправильно распознанных символов, а также средней оценки распознавания.

Показано, что дискретизацию можно представить в виде комбинации затемнения и засветления. Интегральные характеристики модифицированной базы и величины, определяющие качество распознавания, ведут себя согласованно: отклонение между среднестатистическими растрами / полиномиальными векторами баз обучения и распознавания, количество нераспознанных символов и средняя оценка распознавания при дискретизации лежат в диапазоне между соответствующими результатами для затемнения и засветления (или близки к этим результатам).

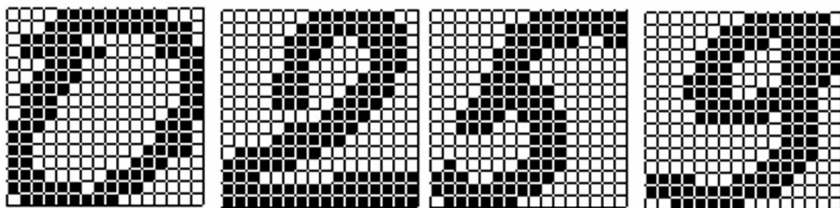


Рис. 1. Образы 16 × 16 рукопечатных цифр



Рис. 2. «Prb_t – Prb_s»

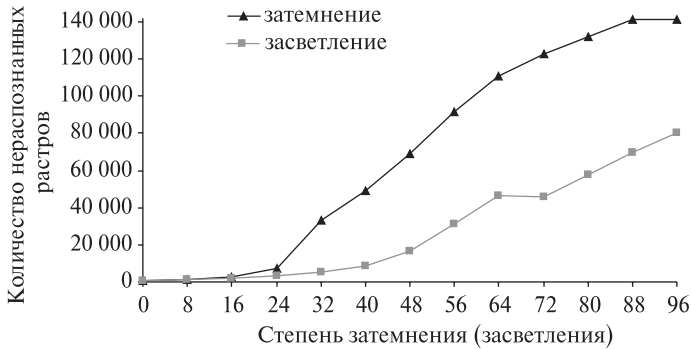


Рис. 3. «mis_t – mis_s»

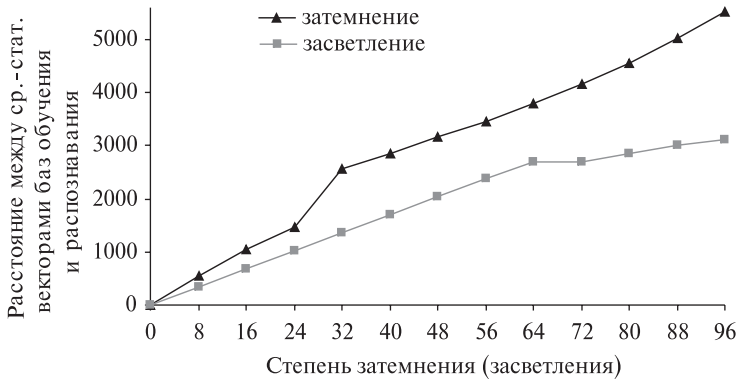


Рис. 4. «xx_t – xx_s»

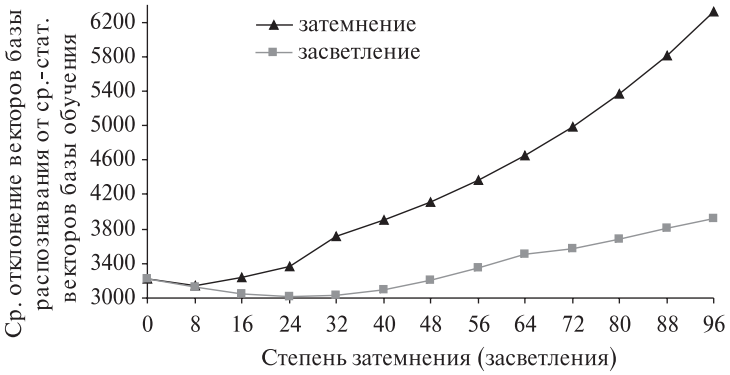


Рис. 5. «dx_t – dx_s»

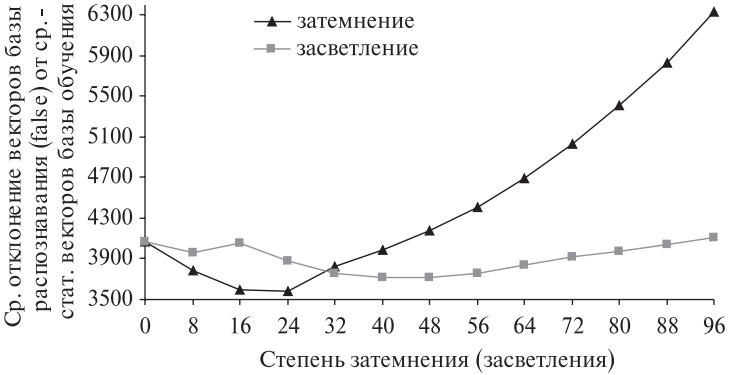


Рис. 6. «dx_{tb} – dx_{sb}»

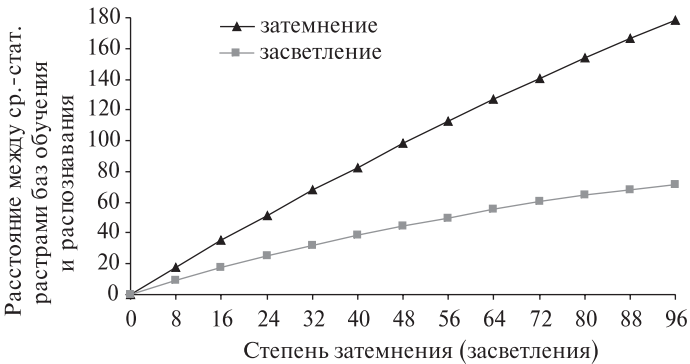


Рис. 7. «vv_t – vv_s»

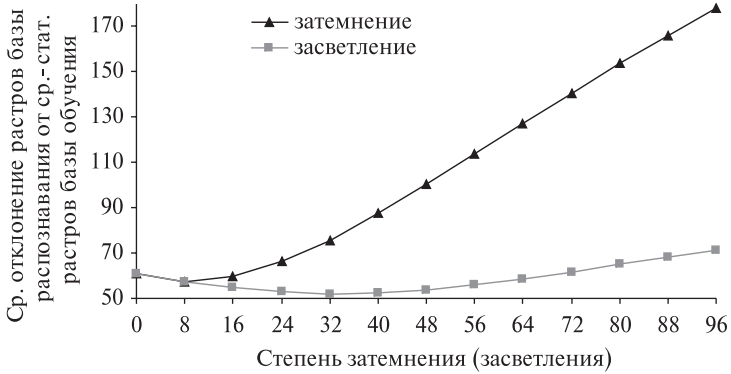


Рис. 8. «dv_t – dv_s»

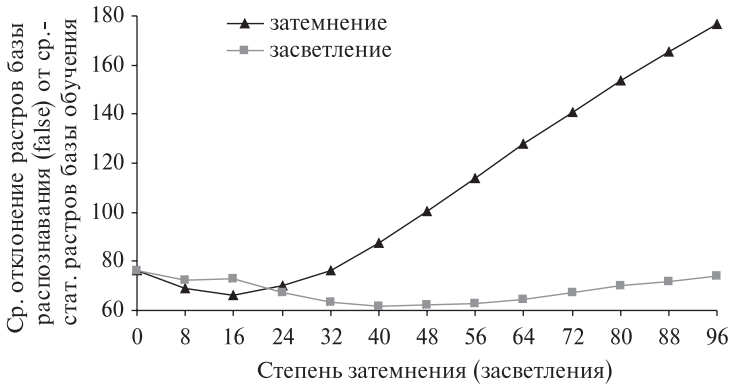


Рис. 9. «dv_{tb} – dv_{sb}»

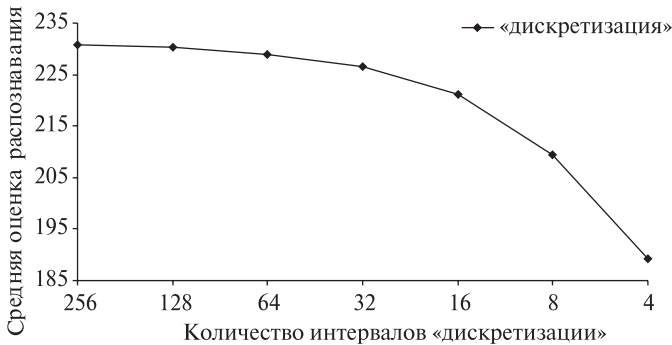


Рис. 10. «Prb»

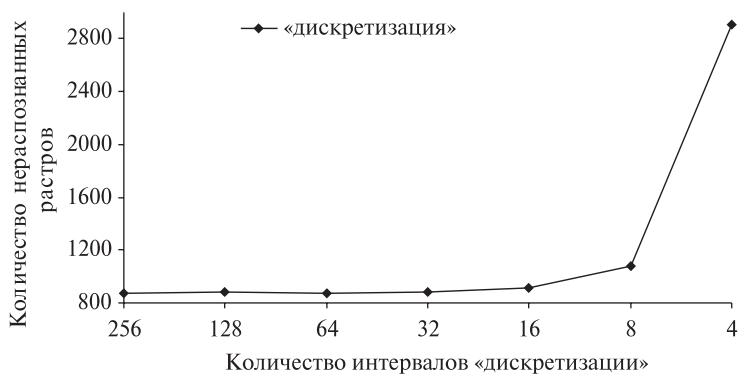


Рис. 11. «mis»

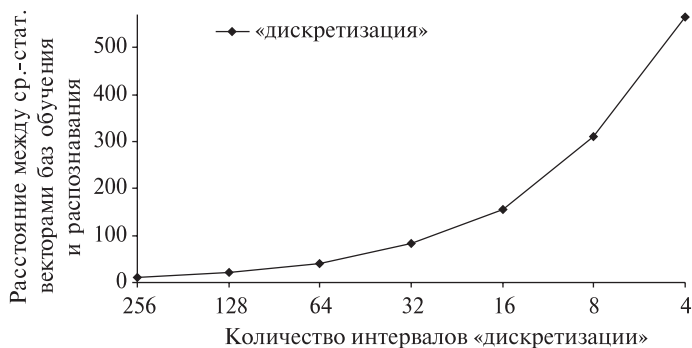


Рис. 12. «xx»

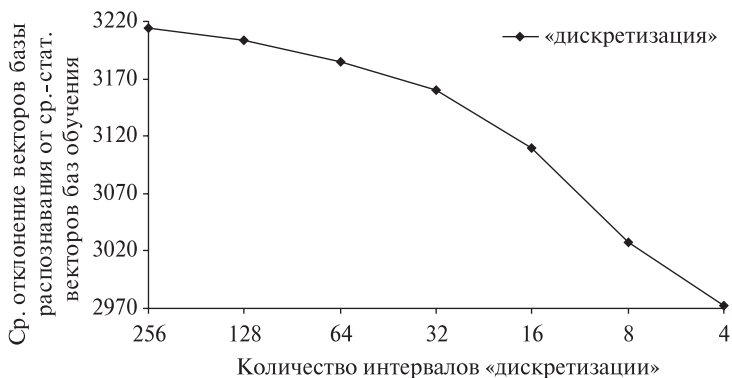


Рис. 13. «dx»

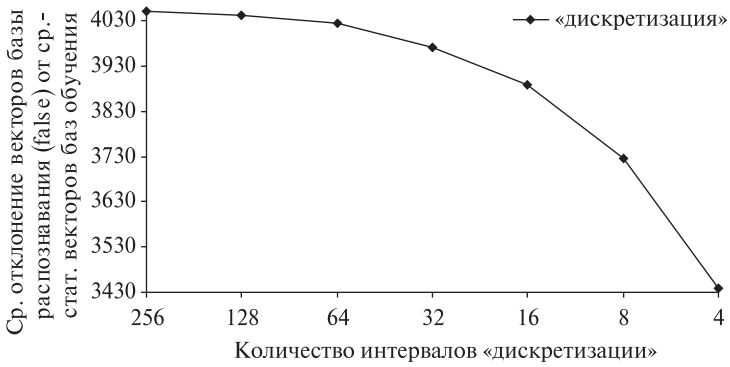


Рис. 14. «dxb»

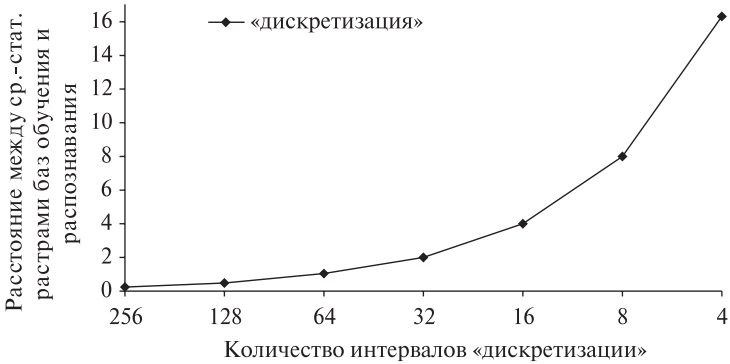


Рис. 15. «vv»

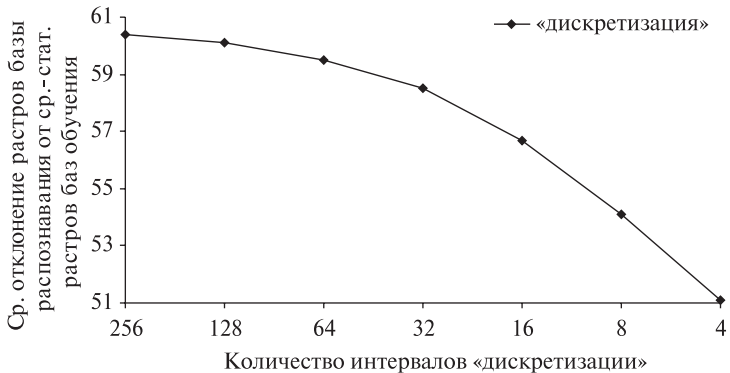


Рис. 16. «dv»

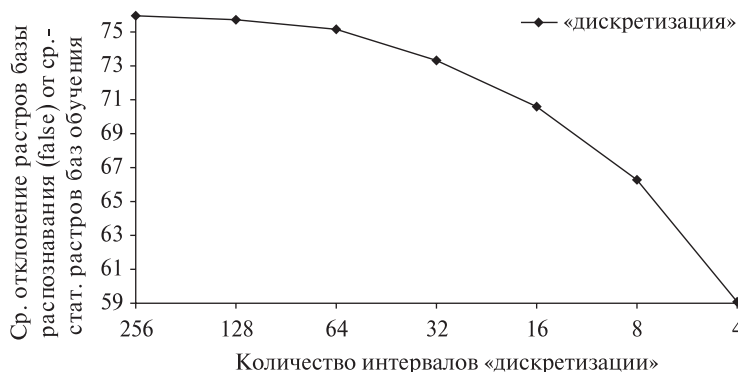


Рис. 17. «dwb»

Литература

1. Гавриков М. Б., Пестрякова Н. В., Усков А. В., Фарсобина В. В. О некоторых свойствах метода распознавания символов, основанного на полиномиальной регрессии // Препринт ИПМ РАН. М., 2007. № 69. 20 с.
2. Гавриков М. Б., Пестрякова Н. В., Усков А. В., Фарсобина В. В. О некоторых свойствах оценки метода распознавания символов, основанного на полиномиальной регрессии // Препринт ИПМ РАН. М., 2008. № 7. 28 с.
3. Гавриков М. Б., Пестрякова Н. В., Усков А. В., Фарсобина В. В. Анализ метода распознавания символов, основанного на полиномиальной регрессии // Препринт ИПМ РАН. М., 2008. № 25. 28 с.
4. Гавриков М. Б., Пестрякова Н. В., Усков А. В., Фарсобина В. В. Оценка распознавания символов для метода, основанного на полиномиальной регрессии // Труды ИСА РАН / Под ред. чл.-корр. РАН, проф. В. Л. Арлазарова и д. т. н., проф. Н. Е. Емельянова. М.: Книжный дом «Либроком»/URSS, 2008. С. 194–219.