



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

О распознавании
модифицированной
обучающей базы. (Часть 2)

Рекомендуемая форма библиографической ссылки: О распознавании модифицированной обучающей базы. (Часть 2) / М.Б.Гавриков [и др.] // Препринты ИПМ им. М.В.Келдыша. 2010. № 13. 16 с. URL: <http://library.keldysh.ru/preprint.asp?id=2010-13>

Введение

Настоящая работа – очередная в серии [1 – 9, 20], посвященной методу распознавания символов, основанному на регрессионном анализе [10 – 19]. Она является непосредственным продолжением [20], поэтому в них применяется сквозная нумерация разделов, таблиц и рисунков.

В качестве базы распознавания использовалась база обучения, модифицированная с применением модели «дискретизации». Проведено сопоставление с результатами работы [20], в которой анализировались модели «затемнения» и «засветления».

При увеличении степени искажения исходной базы найдены закономерности в изменении оценки распознавания (средней по всем верно распознанным изображениям), а также числа неправильно распознанных образов. Выявлены особенности корреляции в поведении указанных величин и основанных на растрах и полиномиальных векторах характеристик базы распознавания, осредненных по совокупности изображений символов.

Выполнен сравнительный анализ зависимостей, полученных для печатных и рукопечатных цифр.

3. Модель «дискретизации»

3.1. 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, как и раньше, будем считать n неотрицательным, но не целым, как в [20], а действительным. От этого не возникает никаких принципиальных расхождений с изложенным в [20], поскольку там тоже можно было считать n действительным, и только для простоты использовались целые значения.

В рамках модели «дискретизации» в каждом пикселе раstra 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: $|(k + 1/2) - F| / N_{\text{дискр}} \leq 1/(2 \cdot N_{\text{дискр}})$. Сле-

довательно, $0,01 \cdot n \leq 1/(2 \cdot N_{\text{дискр}})$. Окончательный результат: $n \leq 50/N_{\text{дискр}}$. В Таблице 3.11 «Дискретизация (рукопечатные цифры)» представлены значения величины $f_{\text{дискр}} = 50 / N_{\text{дискр}}$ в зависимости от $N_{\text{дискр}}$.

3.2. Анализ табличных и графических данных. Согласованность характеристик качества распознавания для дискретизации с соответствующими величинами для затемнения и засветления. В Таблице 7 «Дискретизация (рукопечатные цифры)» и Таблице 8 «Дискретизация (печатные цифры)» 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 [20]. В Таблицах 7, 8 также приводятся число нераспознанных символов mis и средняя оценка распознавания Prb .

Таблица 7 «Дискретизация (рукопечатные цифры)»

	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
mis%	0,50	0,50	0,50	0,50	0,52	0,62	1,66
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

Очевидно, что для произвольного символа уменьшение $N_{\text{дискр}}$ должно сопровождаться увеличением расстояния между среднестатистическими растрами в базах обучения и распознавания vv (Рис.16). Для среднестатистических полиномиальных векторов xx сохраняется указанная закономерность (Рис.13а,б).

Заметим, что темпы роста vv и xx , незначительные вблизи $N_{\text{дискр}} = 256$, нарастают с уменьшением $N_{\text{дискр}}$ – графики являются вогнутыми.

Таблица 8 «Дискретизация (печатные цифры)»

	256	128	64	32	16	8	4
dvw	50,9	50,6	50,1	49,2	47,7	45,9	45,0
dv	50,9	50,6	50,1	49,2	47,7	45,9	45,0
dvb	84,7	84,4	83,7	82,6	80,6	74,0	68,8
dxw	189,0	188,1	186,4	183,2	177,7	170,2	163,0
dx	189,1	188,2	186,4	183,3	177,7	170,3	163,1
dxb	286,5	285,6	283,7	280,3	274,1	255,8	235,5
vv	0,29	0,57	1,2	2,3	4,6	9,3	18,9
xx	1,0	2,0	4,1	8,1	16,4	32,3	64,6
mis	4	4	4	4	4	5	6
mis%	0,07	0,07	0,07	0,07	0,07	0,09	0,11
Prb	234,7	234,0	232,6	229,6	222,7	207,6	176,2

Особенность Рис. 13 – 18 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 $\frac{1}{2}$. 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 пологими и сходными с полученными при затемнении / засветлении.

Дальнейший анализ проводится для шкалы, изображенной на Рис. 13 – 18. 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.16) и xx (Рис.13а,б) при уменьшении $N_{\text{дискр}}$ соответствует уменьшению средней оценки распознавания Prb (Рис.11), причем темпы убывания последней также увеличиваются с уменьшением $N_{\text{дискр}}$. Следует отметить, что и темпы роста vv , и темпы падения Prb для печатных цифр всюду выше, чем для рукопечатных. При этом количество (доля) неправильно распознанных символов mis ($mis\%$) для рукопечатных цифр (Рис.12а,в) 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, а для печатных mis ($mis\%$) неизменно при числе отрезков дискретизации от 256 до 16, а далее нарастает с постоянной скоростью (Рис.12б,в). На всем изучаемом интервале для рукопечатных цифр и доля неправильно распознанных символов, и темпы ее роста больше, чем для печатных.

Отметим, что и для печатных, и для рукопечатных цифр указанные в предыдущем абзаце графики изменения величин при дискретизации с уменьшением $N_{\text{дискр}}$ ($N_{\text{дискр}} = 256, 128, 64, 32, 16, 8, 4$) качественно соответствуют аналогичным для затемнения и засветления с диапазоном степеней n : $0 \leq n \leq 50/4$.

Для рукопечатных цифр 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 Таблице 1 «Затемнение и засветление (рукопечатные цифры)» в [20] видно, что при $n = 1$ для засветления также зафиксирован близкий по значению минимум mis .

Нетрудно проверить, что и при печатном, и при рукопечатном написании 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 Таблице 7 «Дискретизация (рукопечатные цифры)» и Таблице 8 «Дискретизация (печатные цифры)» приведены среднее отклонение растров базы распознавания от среднестатистического растра базы обучения (Рис.17, 18), а также среднее отклонение полиномиальных векторов базы распознавания от среднестатистического вектора базы обучения (Рис.14а,б – 15а,б). Эти величины в Таблицах 7, 8 даны без учета правильности распознавания (Рис.14а,б, Рис.17), для правильно распознанных (аналогично Рис.14а,б, Рис.17) и для неправильно распознанных символов (Рис.15а,б, Рис.18). Используемые обозначения представлены в табличном виде:

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

Расчеты показали, что и для печатных, и для рукопечатных символов при уменьшении $N_{\text{дискр}}$ расстояния dvb , dvw , dv , dxb , dxw , 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; значение dvb существенно больше, чем dv и dvw (аналогично затемнению и засветлению для $0 \leq n \leq 12,5$); из неравенства $dvb >$

dv_w следует неравенство $dv > dv_w$. Кроме того, для всех исследуемых $N_{\text{дискр}}$ выполняются следующие неравенства: $dv_{w_{\text{печ}}} < dv_{w_{\text{рук}}}$, $dv_{\text{печ}} < dv_{\text{рук}}$, но $dv_{b_{\text{печ}}} > dv_{b_{\text{рук}}}$. Заметим, что при затемнении / засветлении для $0 \leq n \leq 12,5$ имеют место более строгие соотношения:

$$\max_{0 \leq n \leq 12,5} (dvs_w, dvt_w)_{\text{печ}} < \min_{0 \leq n \leq 12,5} (dvs_w, dvt_w)_{\text{рук}}$$

$$\max_{0 \leq n \leq 12,5} (dvs, dvt)_{\text{печ}} < \min_{0 \leq n \leq 12,5} (dvs, dvt)_{\text{рук}}$$

$$\min_{0 \leq n \leq 12,5} (dvs_b, dvt_b)_{\text{печ}} > \max_{0 \leq n \leq 12,5} (dvs_b, dvt_b)_{\text{рук}}$$

В терминах полиномиальных векторов также для обоих типов написания величины dx и dx_w отличаются очень мало; значение dx_b существенно больше, чем dx и dx_w (аналогично затемнению и засветлению для $0 \leq n \leq 12,5$); из неравенства $dx_b > dx_w$ следует неравенство $dx > dx_w$.

3.3. Выводы. Для обоих типов написания дискретизацию можно представить в виде комбинации затемнения и засветления. Интегральные характеристики модифицированной базы и величины, определяющие качество распознавания, ведут себя согласованно: отклонение между среднестатистическими расстрами / 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).

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 стр.
- [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 стр.

- [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 стр.
- [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 стр.
- [6] *Гавриков М.Б., Пестрякова Н. В., Усков А.В., Фарсобина В.В.* "Анализ метода распознавания символов, основанного на полиномиальной регрессии ", //Препринт ИПМатем. РАН, М., 2008, №25, 28 стр.
- [7] *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. Арлазарова В.Л. и д.т.н., проф. Емельянова Н.Е. М., «ЛИБРОКОМ», 2008, с. 194-219.
- [8] *Гавриков М.Б., Пестрякова Н. В., Усков А.В., Фарсобина В.В.* "О средне-статистических растрах и векторах метода распознавания символов, основанного на полиномиальной регрессии", // Труды ИСА РАН под ред. чл.-корр. РАН, проф. Арлазарова В.Л. и д.т.н., проф. Емельянова Н.Е. М., «ЛИБРОКОМ», 2008, с. 220-232.
- [9] *Гавриков М.Б., Пестрякова Н. В., Усков А.В., Фарсобина В.В.* "Зависимость точности и оценки распознавания от степени различия между базами обучения и распознавания", // Труды ИСА РАН под ред. чл.-корр. РАН, проф. Арлазарова В.Л. и д.т.н., проф. Емельянова Н.Е. М., «ЛИБРОКОМ», 2008, с.233-240.
- [10] *Sebestyen G.S.* Decision Making Processes in Pattern Recognition, MacMillan, New York, 1962.
- [11] *Nilson N. J.* Learning Machines, McGraw-Hill, New York, 1965.

- [12] *Schürmann J.* Polynomklassifikatoren, Oldenbourg, München, 1977.
- [13] *Schürmann J.* Pattern Classification, John Wiley&Sons, Inc., 1996.
- [14] *Albert A.E. and Gardner L.A.* Stochastic Approximation and Nonlinear Regression // Research Monograph 42. MIT Press, Cambridge, MA, 1966.
- [15] *Becker D. and Schürmann J.* Zur verstärkten Berücksichtigung schlecht erkennbarer Zeichen in der Lernstichprobe // Wissenschaftliche Berichte AEG-Telefunken **45**, 1972, pp. 97 – 105.
- [16] *Pao Y.-H.* The Functional Link Net: Basis for an Integrated Neural-Net Computing Environment // in Yoh-Han Pao (ed.) Adaptive Pattern Recognition and Neural Networks, Addison-Wesley, Reading, MA, 1989, pp. 197-222.
- [17] *Franke J.* On the Functional Classifier, in Association Francaise pour la Cybernetique Economique et Technique (AFCET), Paris // Proceedings of the First International Conference on Document Analysis and Recognition, St. Malo, 1991, pp.481-489.
- [18] *Дж.Себер.* Линейный регрессионный анализ. М.:”Мир”, 1980.
- [19] *Линник Ю.В.* Метод наименьших квадратов и основы математико - статистической теории обработки наблюдений. М.:”Физматлит”, 1958.
- [20] *Гавриков М.Б., Пестрякова Н. В., Усков А.В., Фарсобина В.В.* "О распознавании модифицированной обучающей базы (Часть 1)", //Препринт ИПМатем. РАН, М., 2010, 28 стр. (в печати).

Рисунки

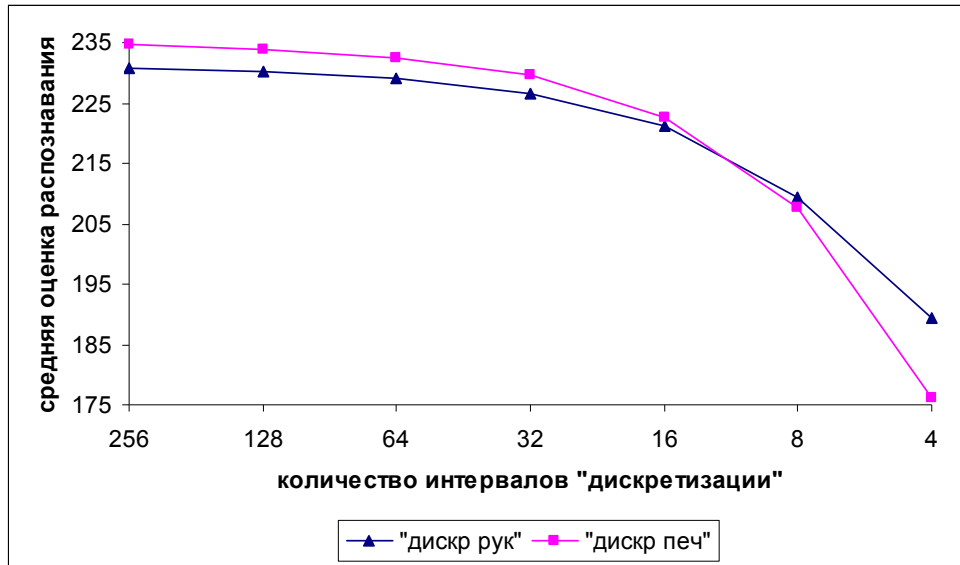
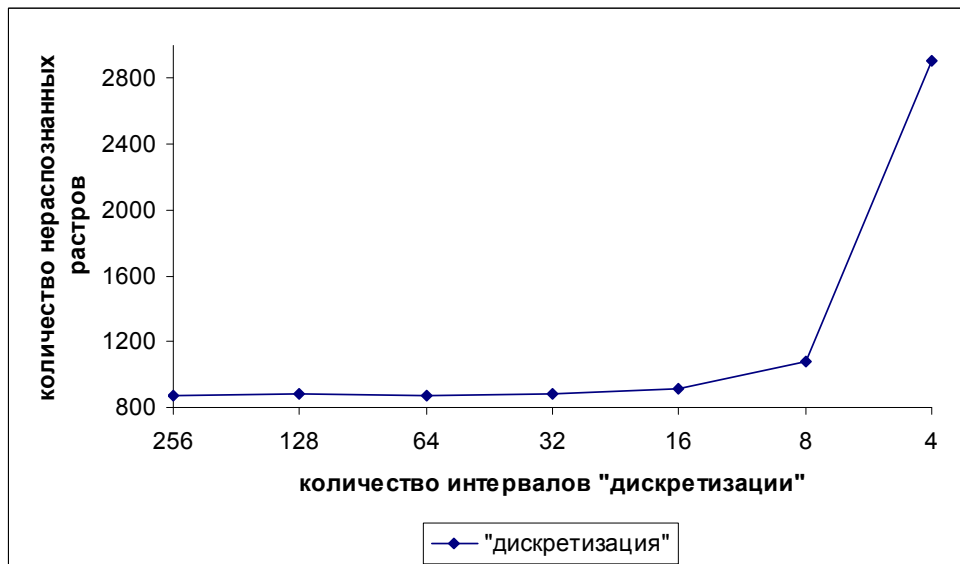
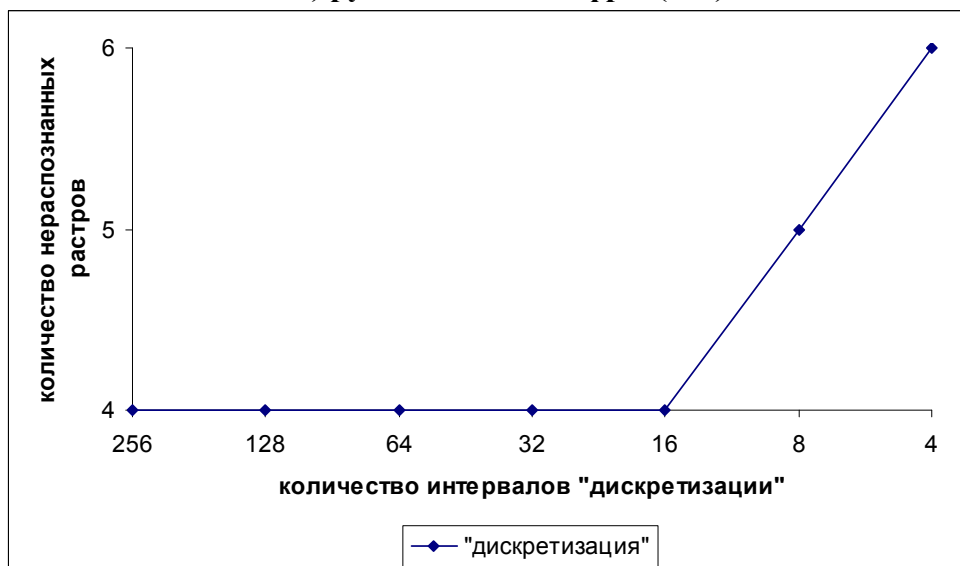


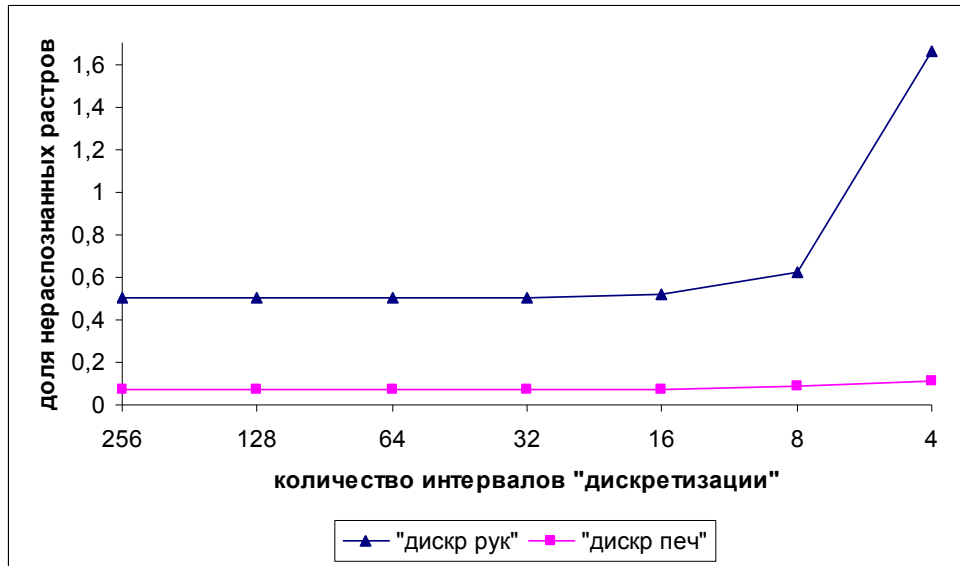
Рис. 11 «рукопечатные и печатные цифры (Prb)»



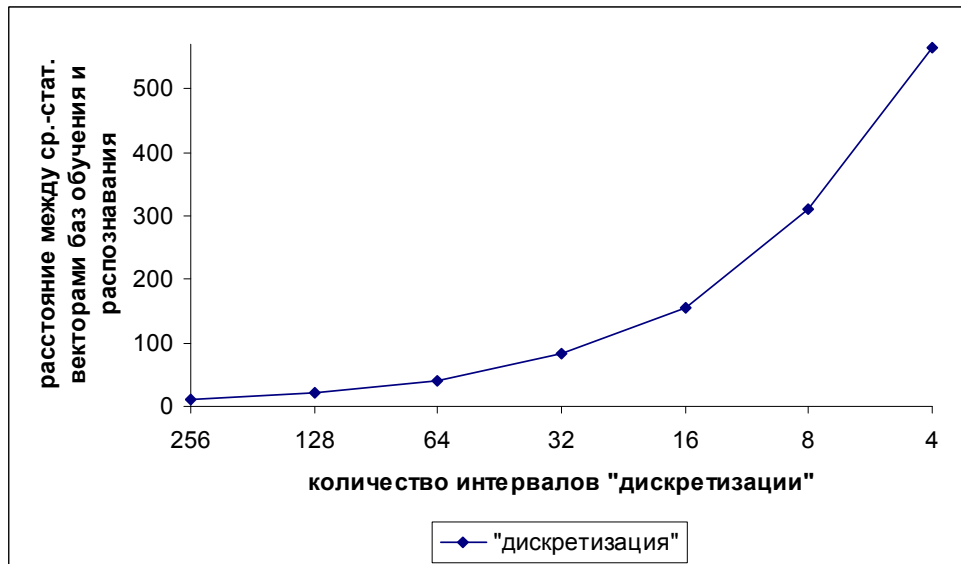
а) рукопечатные цифры (mis)



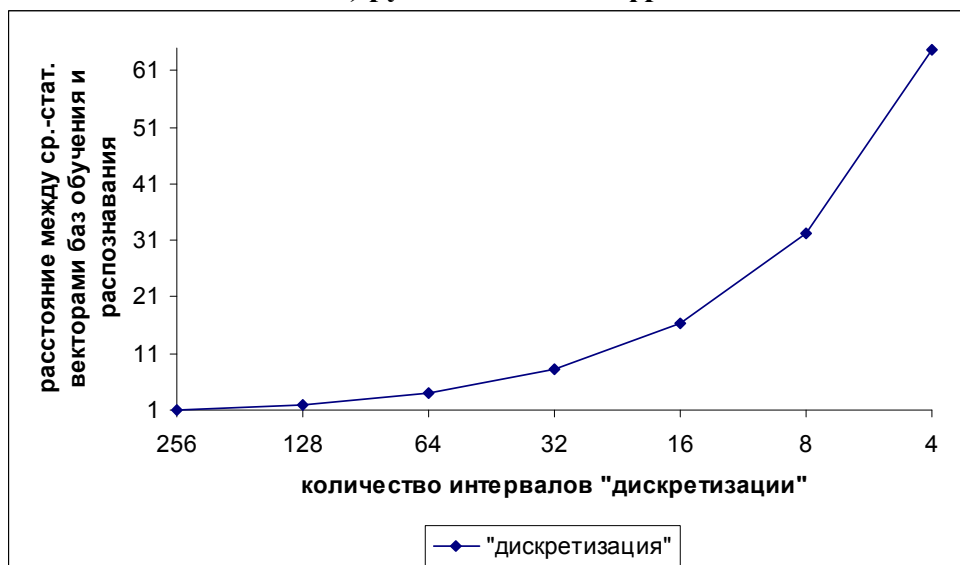
б) печатные цифры (mis)



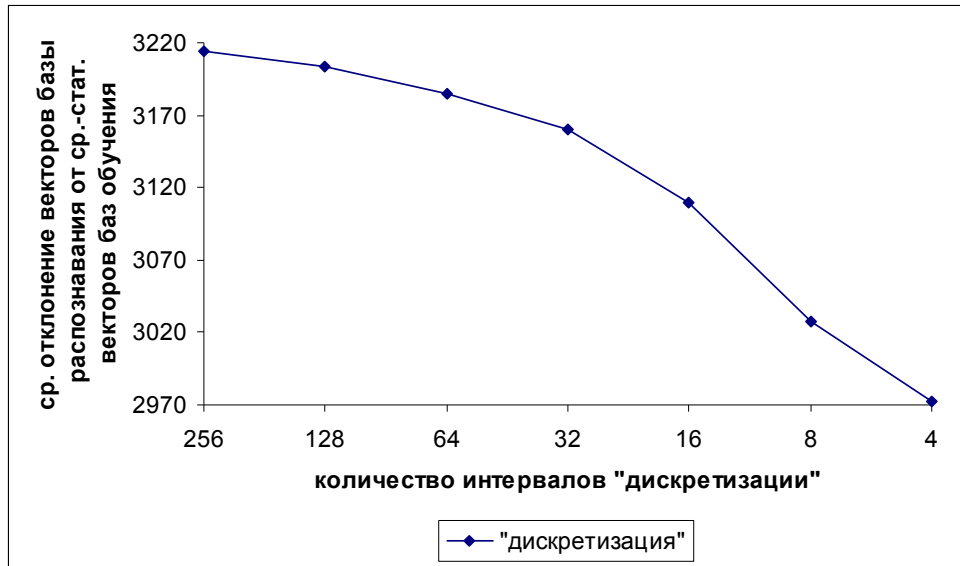
в) рукопечатные и печатные цифры (mis%)
Рис. 12 «mis, mis%»



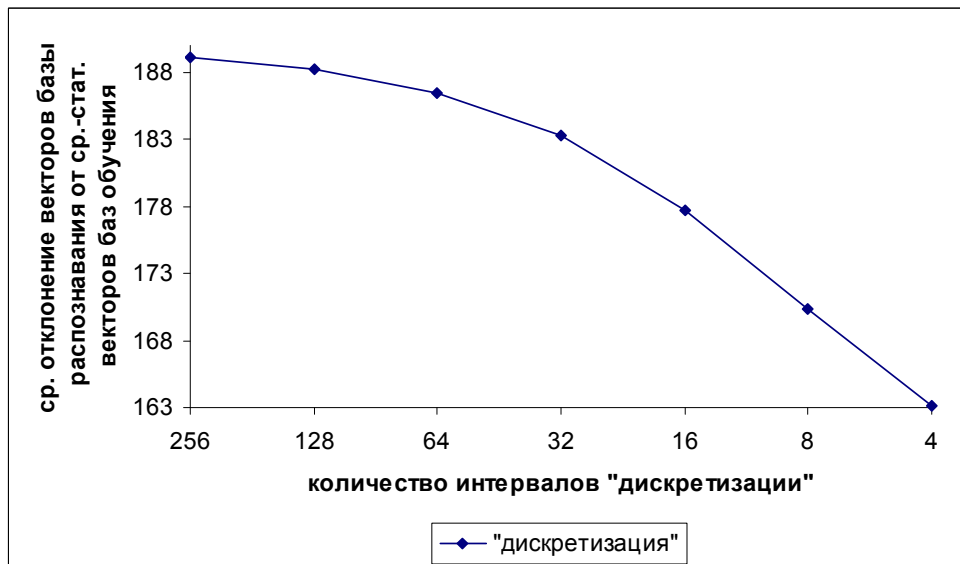
а) рукопечатные цифры



б) печатные цифры
Рис. 13 «xx»

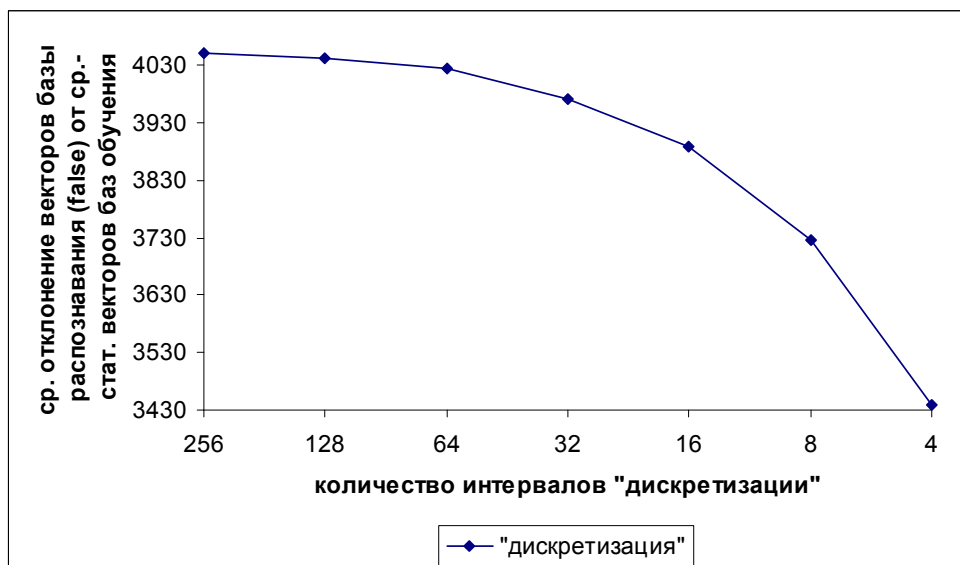


а) рукопечатные цифры

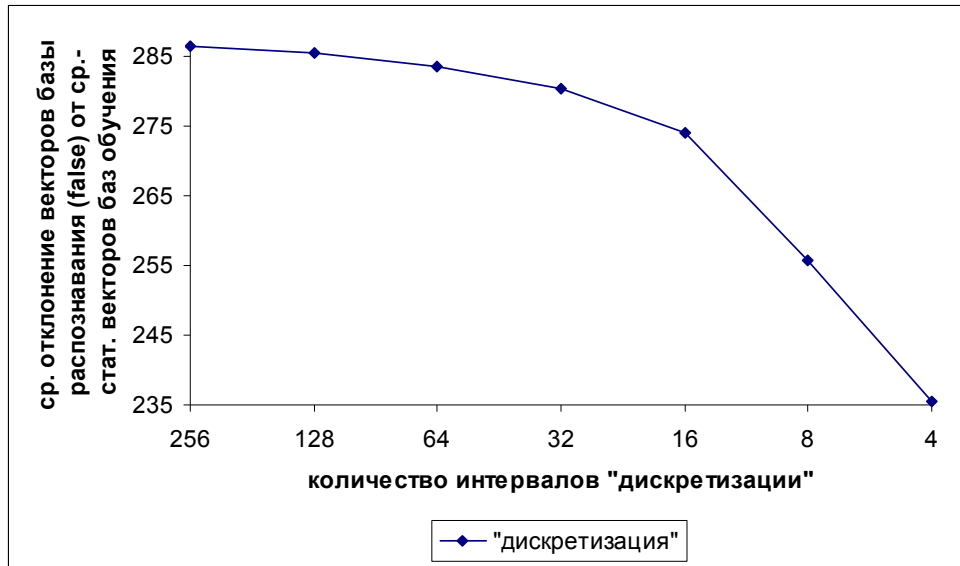


б) печатные цифры

Рис. 14 «dx»



а) рукопечатные цифры



б) печатные цифры
Рис. 15 «dxb»

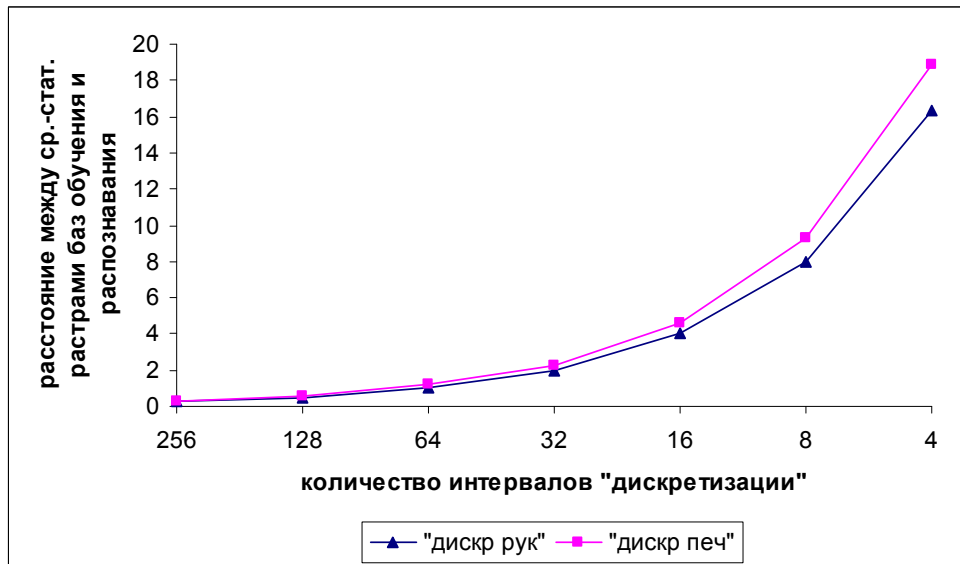


Рис. 16 «рукопечатные и печатные цифры (vv)»

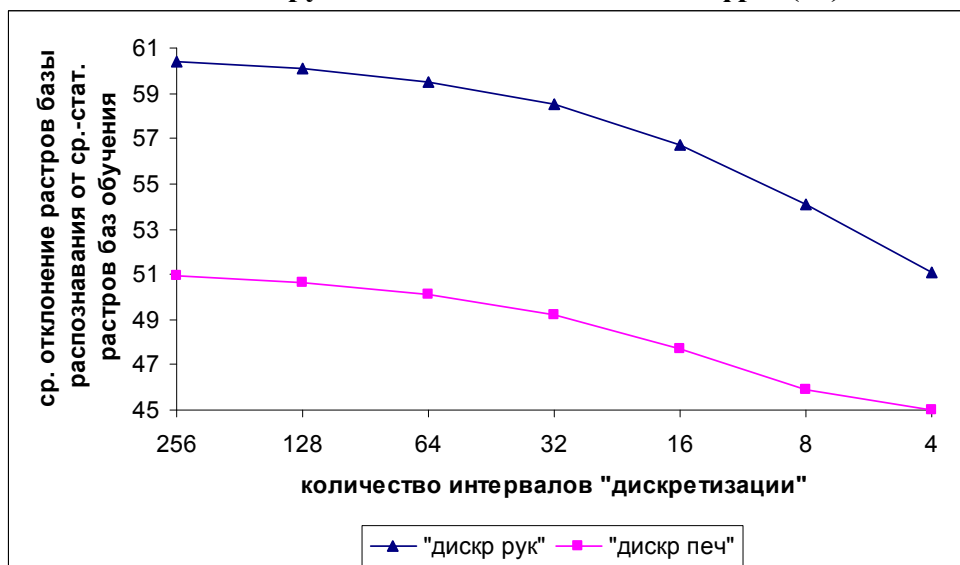


Рис. 17 «рукопечатные и печатные цифры (dv)»

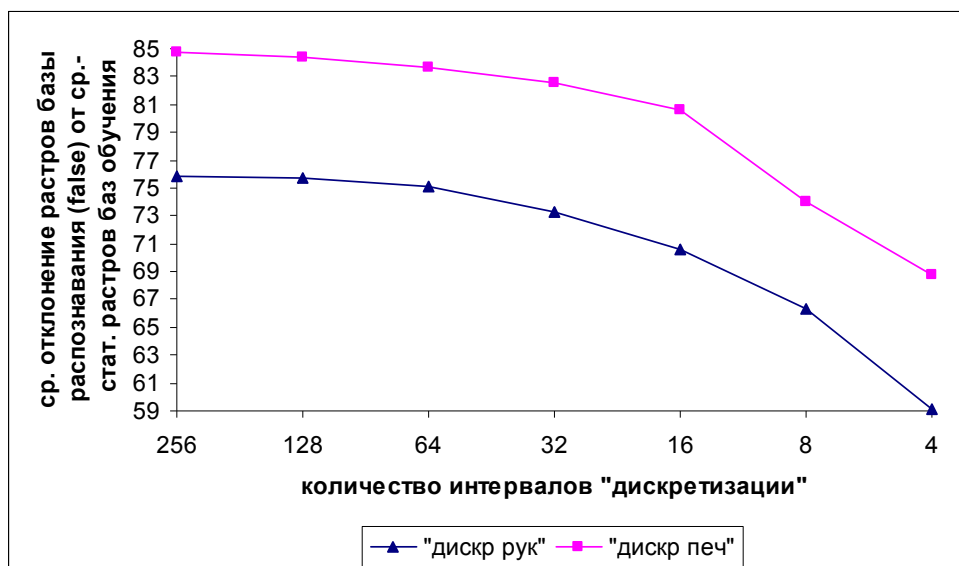


Рис. 18 «рукопечатные и печатные цифры (dvv)»