УДК 681.327.12

Е. В. Волченко

МОДИФИЦИРОВАННЫЙ МЕТОД ПОТЕНЦИАЛЬНЫХ ФУНКЦИЙ

1. Введение

Задача выбора решающего правила классификации является одним из пяти этапов построения системы автоматического распознавания [1]. Она состоит в нахождении (построении) численного выражения или алгоритма, результатом выполнения которого будет номер класса, к которому будет отнесен распознаваемый объект. В зависимости от объема априорной информации, доступной разработчику системы на этапе ее проектирования, системы распознавания делятся на три вида, для каждого из которых существуют свои особенности выбора решающего правила классификации [2]:

- системы без обучения (в качестве решающего правила используется критерий Байеса и его модификации);
- системы с учителем (для заданной классифицированной обучающей выборки необходимо найти решающее правило итерационно в процессе обучения);
- самообучающиеся системы (для заданной неклассифицированной выборки необходимо выделить классы объектов и построить алгоритм классификации во время самообучения).

Как известно, каждый объект в системе распознавания описывается набором признаков, количество которых может быть достаточно велико. Если каждому признаку в многомерном пространстве поставить в соответствие координатную ось, то каждому объекту будет соответствовать точка в полученном пространстве признаков. Объекты, имеющие то или иное сходство, на основе которого они относятся к одному классу в пространстве признаков, локализуются в определенные области. Всем распознаваемым классам в пространстве признаков соответствует несколько областей, поэтому решить задачу распознавания на основе обучающей выборки означает установить поверхность (гиперплоскость), которая разделяет эти области [3].

Для решения задачи распознавания в обучающихся системах существует несколько групп алгоритмов, отличающихся способом построения решающей функции (правила) [4]. К первой группе относятся методы, с помощью которых строится некоторый функционал на основе концепции допустимых преобразований. Основной проблемой таких методов являются слабые адаптивные способности. Вторая группа методов характеризуется тем, что решающее правило находится рекуррентно в процессе обучения. Различие между алгоритмами данной группы состоит в выборе аппроксимирующей функции, вида экстремизируемого функционала и способа экстремизации этого функционала. Наиболее известными алгоритмами этой группы являются метод секущих гиперплоскостей, метод потенциальных функций и метод группового учета аргументов. К третьей группе относятся системы персептронного типа и методы лингвистического распознавания. Однако эти методы отличаются низкой помехоустойчивостью и требуют наличия значительной априорной информации о распознаваемых объектах.

В данной статье рассматривается метод потенциальных функций, предложенный в [5], как один из наиболее часто используемых при построении сложных систем и к которому могут быть сведены многие алгоритмы обучения, используемые, в том числе, и в нейронных сетях.

При решении практических задач довольно часто приходится иметь дело с большими выборками объектов, описываемых несколькими десятками признаков. Это приводит к существенным временным и ресурсным затратам при хранении и обработке таких выборок. Поэтому очевидна потребность обработки обучающей выборки с целью сокращения ее длины, не уменьшая при этом качества обучения и распознавания [6].

В [7] предлагается два алгоритма уменьшения размера обучающей выборки. Алгоритм STOLP включает в новую выборку только «точки опоры», расстояние до которых от объектов «своего» класса меньше расстояния от объектов «чужого» класса. Существенным недостатком алгоритма является его комбинаторная сложность. При выполнении распознавания по «точкам опоры» предлагается использовать метод ближайшего соседа [8], который рассматривает только один ближайший объект каждого из классов, что может приводить к ошибкам классификации. Алгоритм ДРЭТ (метод «дробящихся эталонов») основан на идее покрытия всего обучающего множества объектов в признаковом пространстве сферами минимального радиуса. Каждая сфера покрывает только объекты одного класса, а ее радиус подбирается во время работы алгоритма. При проведении распознавания объект будет отнесен к тому классу, расстояние до центра ближайшей сферы которого минимально.

В [9] предлагается алгоритм формирования обучающей выборки по алгоритму, сходному с алгоритмом STOLP. Выбор «узловых точек» осуществляется в результате анализа матрицы расстояний.

На наш взгляд, общим недостатком методов, приведенных в [7] и [9], является исключение из обучающей выборки некоторой части объектов без сохранения значений их признаков для построения решающей функции.

Авторы статьи [10] предлагают устранить этот недостаток посредством нахождения центра сосредоточения объектов каждого из классов и замены всех объектов класса на один центральный объект. При проведении такой замены существенно уменьшаются временные и ресурсные характеристики алгоритмов обучения и распознавания, однако при пересечении классов в признаковом пространстве эффективность алгоритма существенно снижается. Процесс распознавания в алгоритме также сводится к нахождению класса, расстояние до центрального объекта которого минимально.

В данной статье предлагается иной подход к построению решающего правила методом потенциальных функций. Он заключается в том, что группа близкорасположенных объектов одного класса заменяется одним мета-объектом. Из мета-объектов формируется новая обучающая выборка, которая является исходными данными для модифицированного метода потенциальных функций. Предложенный подход позволяет существенно уменьшить время обучения и распознавания, сократить объем памяти для хранения обучающей выборки, однако при этом сохранить эффективность распознавания.

2. Постановка задачи

Пусть задана некоторая обучающая классифицированная выборка объектов $W=\{\vec{W}_1, \vec{W}_2, ..., \vec{W}_s\}$, каждый из которых относится к классу V_1 или V_2. Каждый объект задан набором своих признаков $\vec{W}_i=\{w_{i1}, w_{i2}, ..., w_{in}\}$, где w_{ij} — j-й признак i-го объекта обучающей выборки. Необходимо найти такую разделяющую функцию $y=f(W)$, которая по своему знаку определяет две области в n-мерном пространстве признаков:

$$\operatorname{sign} y = \operatorname{sign} f(W) = \begin{cases} +1, \textit{если } \vec{W}' \in V_1; \\ -1, \textit{если } \vec{W}' \in V_2, \end{cases}$$

где $\vec{W}'$ — распознаваемый объект.

Задача нахождения такой гиперплоскости методом потенциальных функций сводится к рекуррентному построению функции вида

$$U(\vec{W}', W_k) = \sum_{m=1}^{k} U(\vec{W}', \vec{W}_m),$$

где W_k — множество из k объектов обучающей выборки, включенных в потенциальную функцию; $U(\vec{W}', \vec{W}_m)=\exp(-\alpha \cdot d^2)$, $d^2=\left\|\vec{W}'-\vec{W}_m\right\|$; α — параметр крутизны потенциальной функции, выбираемый эвристически.

Для оценки качества распознавания построенной потенциальной функции [5] введем меру $N(U(\vec{W}', W_k))$, равную количеству неверно классифицированных объектов обучающей выборки потенциальной функцией $U(\vec{W}', W_k)$.

В качестве эталонного значения N_0 меры $N(U(\vec{W}', W_k))$ для дальнейшей оценки произвольной потенциальной функции примем количество неверно классифицированных объектов функцией, включающей в себя все объекты обучающей выборки $N(U(\vec{W}', W_k))$.

Очевидно, что время выполнения классификации прямо пропорционально числу k. Поэтому при решении задачи нахождения оптимальной разделяющей функции необходимо минимизировать это число без потери качества распознавания, т. е. должны выполняться следующие условия:

$$N(U(\vec{W}', W_k)) = N_0;\ k \to \min. \qquad (1)$$

3. Описание метода построения мета-выборки и алгоритма построения решающей функции

Для простоты геометрического представления объектов обучающей выборки в виде точек, расположенных в пространстве признаков П, без потери сущности рассматриваемой задачи примем, что каждый объект обучающей выборки описывается двумя некоррелированными признаками w_1 и w_2, а система распознавания имеет только два класса в составе алфавита.

При изучении возможных вариантов расположения объектов двух классов в признаковом пространстве были выделены следующие ситуации:

1) классы в признаковом пространстве линейно разделимы, что позволяет использовать в качестве решающего правила одну плоскость (рис. 1);

2) классы линейно неразделимы, однако при разделении их некоторой плоскостью количество неверно классифицированных объектов обучающей выборки достаточно мало (рис. 2);

3) классы в признаковом пространстве существенно пересекаются, и среди объектов одного класса имеются области, в которых присутствуют только неверно классифицируемые объекты другого класса (рис. 3);

4) классы в признаковом пространстве существенно пересекаются, нет областей, в которых присутствуют только неверно классифицируемые объекты одного из классов (рис. 4).

Анализируя все возможные варианты расположения объектов в признаковом пространстве (рис. 1–4), приходим к выводу, что для всех ситуаций характерно

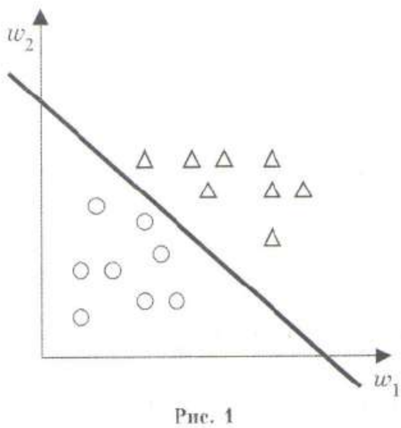


Рис. 1

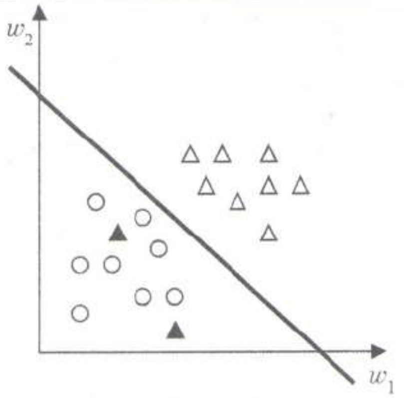


Рис. 2

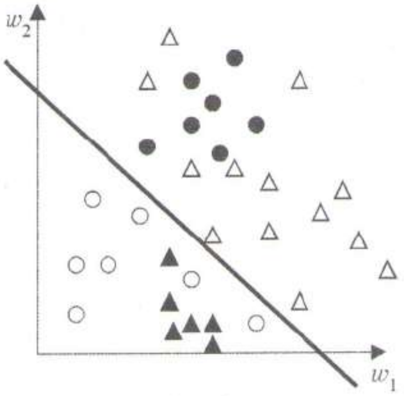


Рис. 3

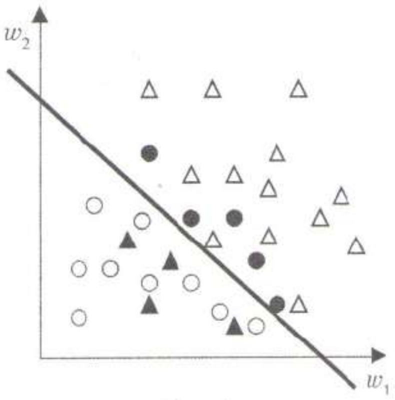


Рис. 4

наличие областей, в которых находятся объекты только одного из классов. Если предположить, что построена решающая функция, которая верно классифицирует все объекты обучающей выборки, то объекты такой области будут классифицироваться одинаково. Очевидно, что в таком случае в обучающей выборке может быть оставлен только один из этих объектов. Однако такой подход не позволит учитывать количество объектов в каждой из групп и их расположение относительно других групп, что может привести к существенному изменению решающей функции, построенной по усеченной выборке по сравнению с исходной выборкой. В связи с этим нами предлагается метод объединения близкорасположенных объектов в один мета-объект и модифицированный метод потенциальных функций, позволяющий учитывать количество объектов, объединенных в мета-объект.

Он состоит в выполнении следующих шагов.

1. Рассчитаем расстояния между всеми парами объектов двух классов

$$r_{ij} = \left\| \overline{W_i} - \overline{W_j} \right\|, \qquad (2)$$

где $\overline{W_i} \in V_1$, $i = \overline{1, |V_1|}$; $\overline{W_j} \in V_2$, $j = \overline{1, |V_2|}$; $|V_l|$ — мощность класса V_l.

2. Выбираем объект f одного из классов, для которого сумма расстояний до всех объектов другого класса, рассчитанная по формуле (2), максимальна:

$$f = \arg \max_{j=\overline{1, |V_1|}} \sum_{j=1}^{|V_2|} r_{ij}.$$

3. Находим объект s другого класса, до которого расстояние от f минимально:

$$s = \arg \min_{j=\overline{1, |V_1|}} r_{fj}.$$

4. Выбираем все объекты того же класса, что и f, расстояние до которых от f меньше, чем до s (рис. 5), и помещаем их во множество W_f

$$W_f = \{\overline{W} \mid \overline{W_i} \in W_f \ (r_{fi} < r_{si})\}.$$

Отметим, что $\overline{W_f} \in W_f$.

5. Заменим W_f мета-объектом $\overline{MW_f}$. При этом будем говорить, что объект $\overline{W_f}$ является включенным в мета-объект $\overline{MW_f}$, если он принадлежит множеству W_f. Значения признаков этого мета-объекта рассчитаем как средние значения признаков всех объектов, в него включенных:

$$w_{f1} = \frac{\sum_{i=1}^{p} w_{i1}}{p}; \; w_{f2} = \frac{\sum_{i=1}^{p} w_{i2}}{p},$$

где $p = |W_f|$ — вес мета-объекта $\overline{MW_f}$.

Таким образом получен мета-объект $\overline{MW_f}$ со значениями признаков $\{w_{f1}, w_{f2}\}$.

6. Удалим из исходной обучающей выборки все объекты, включенные в созданный мета-объект $\overline{MW_f}$.

7. Выполняем пп. 2–5 до тех пор, пока в исходной обучающей выборке не останется ни одного объекта. Сформированную новую обучающую выборку мета-объектов назовем мета-выборкой.

Из описания метода следует, что:

1) один мета-объект может содержать от 1 до объектов исходной обучающей выборки;

2) два мета-объекта не могут содержать один и тот же объект исходной выборки;

3) мета-объект будет относиться к тому же классу, что и все объекты, в него включенные.

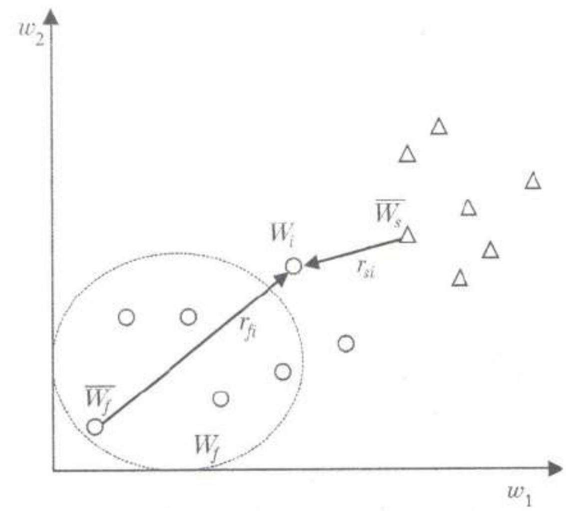


Рис. 5

Для нахождения разделяющей функции по мета-выборке будем использовать метод потенциальных функций. Для учета количества объектов, включенных в каждый из мета-объектов, выполним модификацию этого метода. С этой целью заменим вектор принадлежности объектов потенциальной функции к одному из классов

$$\alpha = \{\alpha_1, ..., \alpha_j, ..., \alpha_k\},$$

где $\alpha_j = \begin{cases} +1, \text{если } \overline{W_j} \in V_1; \\ -1, \text{если } \overline{W_j} \in V_2, \end{cases}$

на вектор

$$\alpha' = \{\alpha'_1, ..., \alpha'_f, ..., \alpha'_{mk}\},$$

где $\alpha'_f = \begin{cases} +p, \text{если } \overline{MW_f} \in V_1; \\ -p, \text{если } \overline{MW_f} \in V_2. \end{cases}$

Введение вектора α' позволяет учитывать совместные значения потенциалов всех объектов, включенных в мета-объекты. Заметим, что если объектам обучающей выборки приписаны веса m, то полагаем, что

$$\alpha'_f = \begin{cases} +\sum_{t=1}^{p} m_t, \text{если } \overline{MW_f} \in V_1; \\ -\sum_{t=1}^{p} m_t, \text{если } \overline{MW_f} \in V_2. \end{cases}$$

4. Результаты экспериментов

Предложенный выше подход к построению решающего правила модифицированным методом потенциальных функций по мета-выборке был реализован программно. Для оценки его эффективности использовалась обучающая выборка, значения признаков объектов которой распределены по нормальному закону, достаточно распространенному в прикладных задачах.

При проведении экспериментов оценивались длина потенциальной функции k и количество неверно классифицированных объектов исходной обучающей выборки $N(U(\bar{W}', W_k))$, описанные в условиях (1).

В каждом из экспериментов были получены три решающие функции:

1) $U(\bar{W}', W_k)$, построенная простым методом потенциальных функций по исходной обучающей выборке;

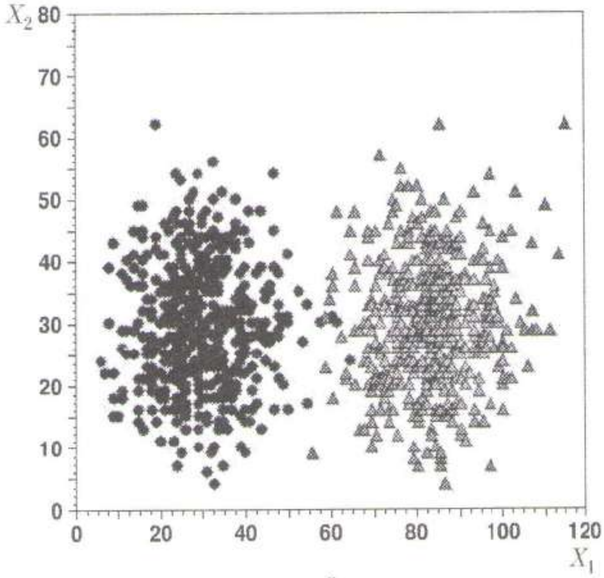
2) $U(\bar{W}', MW_{ms})$, в которую включены все мета-объекты, где ms — количество мета-объектов;

3) $U(\bar{W}', MW_{mk})$, построенная модифицированным методом потенциальных функций по мета-выборке, где mk — количество объектов, включенных в потенциальную функцию, построенную модифицированным методом по мета-выборке.

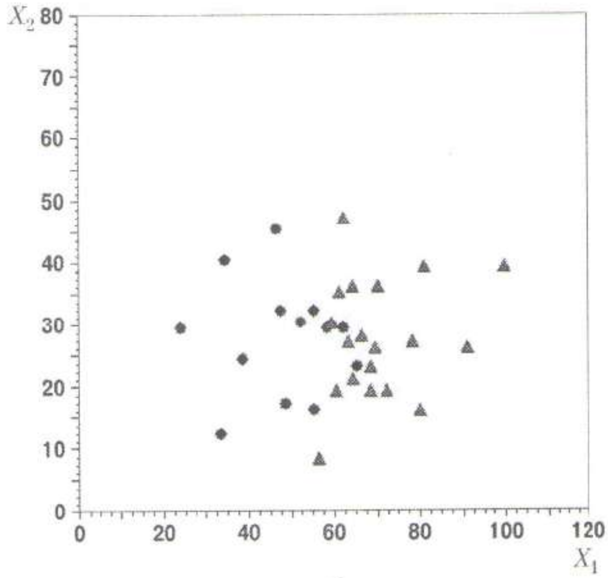
При проведении испытаний использовались обучающие выборки длиной 200...1000 объектов, так как, согласно [11], дальнейшее увеличение объема выборки не улучшает качества получаемого решающего правила. При оценке решающих функций для различной площади пересечения классов рассматривались ситуации, когда классы пересекаются не более чем на 70%, что соответствует данным, используемым в реальных системах. Для непересекающихся классов рассматривалось две ситуации:

- классы расположены близко друг к другу;
- классы существенно удалены друг от друга.

На рис. 6 приведены исходная обучающая выборка (*а*) и полученная по ней мета-выборка (*б*). Анализ результатов объединения объектов в мета-объекты показал, что вес мета-объекта тем больше, чем дальше этот мета-объект находится от границы между классами (области пересечения классов). Было получено, что в зависимости от размера исходной обучающей выборки и площади пересечения классов использование мета-объектов позволяет уменьшить обучающую выборку в 5...300 раз.



а



б

Рис. 6. Исходная обучающая выборка (*а*) и построенная по ней мета-выборка (*б*)

В таблицах 1–3 приведена зависимость длины решающей функции и количества неверно классифицированных объектов обучающей выборки от ее длины для близкорасположенных, но не пересекающихся классов. В таблицах 4–6 приведена зависимость длины решающей функции и количества неверно классифицированных объектов обучающей выборки от площади пересечения классов в признаковом пространстве при размере обучающей выборки 1000 объектов (по 500 объектов каждого из классов). Все значения, приведенные в таблицах, являются средними по 20 проведенным экспериментам.

Анализ приведенных в таблицах 1–6 результатов экспериментов показывает, что для любого размера обучающей выборки и площади пересечения классов длина решающей функции $U(\bar{W}', MW_{mk})$ в среднем на 30 % меньше длины $U(\bar{W}', W_k)$. При этом количество неверных классификаций объектов исходной обучающей выборки $N(U(\bar{W}', MW_{mk}))$ на 10 % меньше $N(U(\bar{W}', W_k))$. Сравнительный анализ $U(\bar{W}', MW_{ms})$ с $U(\bar{W}', W_k)$ и $U(\bar{W}', MW_{mk})$ показывает, что:

- длина $U(\bar{W}', MW_{ms})$ в 2 и 2,5 раза больше длины $U(\bar{W}', W_k)$ и $U(\bar{W}', MW_{mk})$ соответственно;
- $N(U(\bar{W}', MW_{ms}))$ в среднем на 40 % и 30 % меньше $N(U(\bar{W}', MW_{mk}))$ и $N(U(\bar{W}', W_k))$ соответственно.

Таблица 1

Размер обучающей выборки	Метод потенциальных функций, реализованный по исходной выборке	
	k	$N(U(\bar{W}', W_k))$
200	3,62	0,55
400	4,91	0,85
600	5,57	1,12
800	6,34	1,18
1000	7,79	1,25

Таблица 2

Размер обучающей выборки	Метод потенциальных функций, реализованный по мета-выборке	
	mk	$N(U(\bar{W}', MW_{mk}))$
200	2,45	0,47
400	2,65	0,81
600	3,73	1,07
800	4,51	1,08
1000	4,75	1,1

Таблица 3

Размер обучающей выборки	Использование в качестве потенциальной функции всех объектов мета-выборки	
	ms	$N(U(\bar{W}', MW_{ms}))$
200	6,3	0
400	7,3	0
600	9,5	0
800	11,7	0
1000	16,2	0

Таблица 4

Площадь пересечения классов, %	Метод потенциальных функций, реализованный по исходной выборке	
	k	$N(U(\bar{W}', W_k))$
Классы существенно обособлены	3,1	0,5
0	7,8	1,25
10	13,6	3,4
20	23,4	5,75
30	45,6	13,5
40	74,2	30,8
50	130,1	55,4
60	205,7	89,8
70	303,3	161,6

Таблица 5

Площадь пересечения классов, %	Метод потенциальных функций, реализованный по мета-выборке	
	mk	$N(U(\bar{W}', MW_{mk}))$
Классы существенно обособлены	2,1	0,1
0	4,75	1,1
10	8,9	3,1
20	15,5	5,3
30	34,3	12,3
40	59,3	26,6
50	104,2	47,6
60	175,3	87,7
70	263,5	140,4

Таблица 6

Площадь пересечения классов, %	Использование в качестве потенциальной функции всех объектов мета-выборки	
	ms	$N(U(\bar{W}', MW_{ms}))$
Классы существенно обособлены	6,7	0
0	16,2	0
10	32	0,35
20	49,1	0,85
30	95,1	4,7
40	159,7	14,1
50	283,8	32,6
60	434,1	65,3
70	589,7	128,1

Также при проведении экспериментальных исследований были рассчитаны среднеквадратический разброс σ_k длины мета-выборки и среднеквадратический разброс $\sigma_{N(U(\bar{W}', MW_{mk}))}$ количества неверно классифицированных объектов потенциальной функцией, построенной по мета-выборке при различном порядке следования объектов в обучающей выборке (порядок следования объектов в обучающей мета-выборке изменялся посредством «взбалтывания»). Результаты этого исследования для исходных обучающих выборок различных размеров приведены в таблицах 7 и 8. Было получено, что вне зависимости от длины обучающей выборки $\sigma_{N(U(W', MW_{mk}))}$ больше $\sigma_{N(U(W', W_k))}$, т. е. модифицированный метод потенциальных функций более чувствителен к порядку следования мета-объектов в мета-выборке.

Таблица 7

Площадь пересечения классов, %	Метод потенциальных функций, реализованный по исходной выборке	
	σ_k	$\sigma_{N(U(W', W_k))}$
Классы существенно обособлены	0,8	0,44
0	1,67	3,76
10	1,62	2,87
20	2,37	4,28
30	3,04	4,35
40	4,36	9,72
50	5,8	15,8

Таблица 8

Площадь пересечения классов, %	Метод потенциальных функций, реализованный по мета-выборке	
	σ_k	$\sigma_{N(U(W', MW_{mk}))}$
Классы существенно обособлены	0	1,4
0	0,93	5,97
10	1,26	10,33
20	1,85	18,75
30	2,61	23,25
40	4,11	27,19
50	5,09	28,56

5. Выводы

В настоящей работе предложен модифицированный метод потенциальных функций для нахождения решающего правила обучающейся системы распознавания. В качестве обучающей выборки для реализации этого метода было предложено использовать выборку мета-объектов, методика формирования которой также описана в данной статье. Проведенные эксперименты показали преимущества предложенного подхода по сравнению с простым методом потенциальных функций.

Автор благодарен Л. А. Белозерскому и И. С. Грунскому за ряд полезных замечаний, способствовавших улучшению статьи.

Список литературы: 1. *Белозерский Л. А.* Введение в теорию автоматического распознавания.– Киев: Наук. думка. 2005.– 434 с. 2. *Горелик А. Л., Скрипкин В. А.* Методы распознавания: Учеб. пособ. для вузов.– М.: Высшая школа, 1977.– 222 с. 3. *Букатова Л. И.* Эволюционное моделирование: идеи, основы теории, приложения.– М.: Знание, 1981.– 64 с. (Новое в науке и технике. Сер. «Математика, кибернетика»; № 10). 4. *Васильев В. И.* Распознающие системы: Справочник.– Киев: Наук. думка, 1983.– 422 с. 5. *Айзерман М. А., Браверманн Э. М., Розоноэр Л. И.* Метод потенциальных функций в теории обучения машин.– М.: Наука, 1970.– 384 с. 6. *Белозерский Л. А.* Совместимость обучаемого классификатора и систем автоматического распознавания // Искусственный интеллект.– 2003.– № 4.– С. 325–344. 7. *Загоруйко Н. Г.* Прикладные методы анализа знаний и данных.– Новосибирск: Изд-во Ин-та математики, 1999.– 270 с. 8. *Дуда Р., Харт П.* Распознавание образов и анализ сцен.– М.: Мир. 1976. 9. *Дубровин В. И., Субботин С. А., Кривенко В. И., Евченко Л. Н.* Сокращение объема данных в задачах распознавания и диагностики // Тр. VIII Всероссийской конференции «Нейрокомпьютеры и их применение» (Москва, 21–22 марта 2002 г.) / Под редакцией проф. А. И. Галушкина.– М.: Ин-т проблем управления им. В. А. Трапезникова РАН, 2002.– С. 954–963 10. *Дубровин В. И., Корецкий Н. Х., Субботин С. А.* Модифицированный метод потенциальных функций // Складні системи і процеси.– 2002.– № 1.– С. 12–19. 11. *Жук А. В.* Обучаемые классификаторы в статистических испытаниях // Искусственный интеллект.– 2005.– № 1.– С. 168–179.

Поступила в редколлегию 23.04.2006