

РЕГУЛИРУЕМАЯ СЕЛЕКТИВНОСТЬ В МНОГОМОДАЛЬНОМ
РАСПОЗНАВАНИИ ОБРАЗОВ© Моттль В.В.¹, Татарчук А.И.¹, Елисеев А.П.²¹ВЦ РАН, Вавилова 40, Москва, Россия²МФТИ, Институтский переулок 9, г. Долгопрудный, Московская обл., Россия

E-MAIL: vmottl@yandex.ru, aitech@yandex.ru, andreyl@gmail.com

Abstract. The problem of multi-modal pattern recognition is considered under the assumption that the kernel-based approach is applied, which assumed that within each particular modality a kernel function can be specified. The danger of over-fitting makes it necessary to truncate the set of initially available modalities. Two known wrapper-based kernel fusion techniques, Relevance and Support Kernel Machines, offer a toolkit of combining pattern recognition modalities. In this paper, we propose the modifications of the fusion techniques equipped with the ability to preset the desired level of feature-selectivity.

ВВЕДЕНИЕ

В задаче обучения распознаванию образов по прецедентам предполагается, что задан тот или иной конструктивный способ выражения доступной информации об объектах реального мира $\omega \in \Omega$, который принято называть *модальностью* (modality) представления объектов [1, 2]. Модальность может быть выражена в виде признака объекта $x(\omega) \in \mathbb{X}$ с некоторой шкалой измерения \mathbb{X} или в виде меры попарного сходства $\rho(\omega', \omega'') \in \mathbb{R}^+$ между объектами $\omega', \omega'' \in \Omega$.

Интенсивное развитие беспризнаковой методологии распознавания образов [3, 1, 5, 6], основанной на понятии потенциальной функции (kernel) $K(\omega', \omega'') \in \mathbb{R}$, унифицирующей различные представления объектов $\omega', \omega'' \in \Omega$ в виде элементов гипотетического линейного пространства $x(\omega'), x(\omega'') \in \mathbb{X}$ со скалярным произведением $K(\omega', \omega'')$, позволяет использовать фактически весь наработанный арсенал линейных методов анализа данных.

Стремление обеспечить требуемое качество распознавания, недостижимое на основе какой-либо одной модальности, привело к появлению огромного разнообразия способов представления объектов ($x_i(\omega) \in \mathbb{X}_i$, $i = 1, \dots, n$) и созданию *многомодальных систем распознавания*, комбинирующих сразу все доступные модальности в единой процедуре распознавания $y(x_1(\omega), \dots, x_n(\omega)) : \mathbb{X}_1 \times \dots \times \mathbb{X}_n \rightarrow \{-1, +1\}$.

Эффект переобучения приводит к проблеме сокращения исходного множества модальностей или отбора признаков (modality or feature selection). Принято разделять все методы отбора признаков на *фильтры* (filters) и *встроенные методы* (wrappers) [7]. Фильтры применяются на множестве всех модальностей до обучения независимо от метода дальнейшей классификации, а встроенные методы существенно зависят от его специфики. На практике так же принято дополнительно различать *непрерывные методы* (continues methods), в которых признакам приписываются некоторые веса, и *дискретные методы* (binary methods), результатом работы которых является строго выбранное подмножество признаков.

В данной работе рассматриваются две модификации метода опорных векторов (SVM) [1], а именно, метод *релевантных потенциальных функций* (Relevance Kernel Machine или RKM) [8, 9] и *метод опорных потенциальных функций* (Support Kernel Machine или SKM) [9, 10], которые являются наиболее яркими представителями, соответственно, непрерывных и дискретных встроенных методов отбора признаков или модальностей, выраженных в терминах потенциальных функций.

Оба метода реализуют байесовское обучение в пространстве параметров разделяющей гиперплоскости по заданной обучающей совокупности при двух разных предположениях об априорных распределениях компонент направляющего вектора гиперплоскости. Однако, не один из методов не обладает механизмом задания требуемого уровня селективности отбора, начиная с обучения сразу на всех доступных признаках и заканчивая выбором только одного признака.

В данной работе предлагается заменить в байесовской модели обучения фиксированные априорные распределения компонент направляющего вектора на параметрические семейства распределений так, чтобы параметр управлял свойством моделей подавлять значения компонент направляющего вектора, ограничивая количество признаков или модальностей, участвующих в распознавании.

Селективность, являясь структурным параметром алгоритма обучения, влияет на сложность обучаемой модели и тем самым представляет собой дополнительный инструмент борьбы с переобучением с целью повышения обобщающей способности при многомодальном распознавании образов. Выбрать подходящий уровень селективности можно одним из стандартных методов, например, по контрольной выборке или скользящим контролем (cross-validation).

Результаты экспериментального исследования на модельных данных наглядно демонстрируют адекватность механизма регулируемой селективности при многомодальном распознавании образов.

1. КВАЗИ-ВЕРОЯТНОСТНЫЙ ПОДХОД

Будем предполагать, что все объекты $\omega \in \Omega$ поделены на два класса $y(\omega) \in \mathbb{Y} = \{-1, +1\}$ и на этом множестве объектов заданы n признаков или модальностей в некоторых шкалах $x_i(\omega) \in \mathbb{X}_i$, $i = 1, \dots, n$. Однако беспризнаковая концепция распознавания образов фактически полностью стирает грань между различными видами представлений объектов и позволяет, для простоты изложения, рассматривать все исходные модальности $x_i(\omega) \in \mathbb{X}_i$ в виде действительных признаков $\mathbb{X}_i = \mathbb{R}$.

Методология обучения распознаванию образов на основе концепции *оптимальной разделяющей гиперплоскости*, вообще говоря, не предусматривает принятия каких-либо предположений о вероятностной модели генеральной совокупности. Такой концепции соответствует качественная модель в виде линейного пространства признаков $(x_1, \dots, x_n) \in \mathbb{R} \times \dots \times \mathbb{R} = \mathbb{R}^n$, в котором объективно существует некоторая гиперплоскость $y(x_1, \dots, x_n | a_1, \dots, a_n, b) = \sum_{i=1}^n a_i x_i + b = 0$ такая, что объекты двух разных классов отображаются, в основном, по разные стороны от нее, без количественного уточнения в какой степени это предположение может нарушаться. При этом выбор *направляющего вектора* $(a_1, \dots, a_n) \in \mathbb{R}^n$ и значения *порога* $b \in \mathbb{R}$

полностью задает некоторую классификацию множества объектов $\omega \in \Omega$ в комбинированном линейном пространстве признаков $(x_1(\omega), \dots, x_n(\omega)) \in \mathbb{R}^n$.

В данной работе для обобщения известных принципов комбинирования потенциальных функций предлагается *квази-вероятностная модель* генеральной совокупности, на основании которой из байесовского подхода к задаче обучения естественным образом вытекает концепция оптимальной разделяющей гиперплоскости. Модель была названа квази-вероятностной поскольку она основана на так называемых *несобственных плотностях распределения* [11], т.е. некоторых функциях, интерпретируемых как плотности распределения, но для которых не существует конечных интегралов по всем переменным.

Будем предполагать существование плотности распределения в пространстве наблюдаемых признаков и скрытых индексов классов $(x_1, \dots, x_n, y) \in \mathbb{R} \times \dots \times \mathbb{R} \times \mathbb{Y}$, а так же, что элементы обучающей совокупности $(X, Y) = \{x_{1j}, \dots, x_{nj}, y_j, j = 1, \dots, N\}$, $x_{ij} = x_{ij}(\omega_j)$, $y_j = y(\omega_j)$, $\omega_j \in \Omega^*$ выбраны независимо.

В качестве модели генеральной совокупности будем рассматривать два параметрических семейства плотностей распределения $\varphi(x_1, \dots, x_n | a_1, \dots, a_n, b, y)$ с $y \in \{-1, +1\}$ связанных с разделяющей гиперплоскостью $\sum_{i=1}^n a_i x_i(\omega) + b \gtrless 0$ в комбинированном признаковом пространстве:

$$\varphi(x_1, \dots, x_n | a_1, \dots, a_n, b, y) = \begin{cases} \text{const}, & y(\sum_{i=1}^n a_i x_i + b) \geq 1, \\ \exp[-c(1 - y(\sum_{i=1}^n a_i x_i + b))], & y(\sum_{i=1}^n a_i x_i + b) < 1. \end{cases}$$

Очевидно, что такие распределения полностью соответствуют качественной модели генеральной совокупности, являясь равномерными при удалении от разделяющей гиперплоскости и лишь количественно определяя через параметр $c > 0$ возможность отображения объектов определенного класса в "чужую" часть пространства признаков.

Будем рассматривать направляющий вектор (a_1, \dots, a_n) разделяющей гиперплоскости $\sum_{i=1}^n a_i x_i + b \gtrless 0$ как случайный вектор с априорной плотностью распределения $\Psi(a_1, \dots, a_n | \mu)$, задаваемой параметром μ . Что же касается величины порога разделяющей гиперплоскости b , то будем считать, что отсутствуют какие-либо априорные предположения о его значении, тогда совместная априорная плотность параметров гиперплоскости $\Psi(a_1, \dots, a_n, b | \mu) \propto \Psi(a_1, \dots, a_n | \mu)$.

В результате, апостериорная плотность распределения параметров разделяющей гиперплоскости с учетом заданной обучающей совокупности

$$\Phi(X | Y, a_1, \dots, a_n, b) = \prod_{j=1}^N \varphi(x_{1j}, \dots, x_{nj} | a_1, \dots, a_n, b, y_j)$$

пропорциональна произведению

$$P(a_1, \dots, a_n, b | X, Y, \mu) \propto \Psi(a_1, \dots, a_n | \mu) \Phi(X | Y, a_1, \dots, a_n, b).$$

Принцип максимизации апостериорной плотности распределения $P(a_1, \dots, a_n, b | X, Y, \mu)$ в пространстве параметров модели (a_1, \dots, a_n, b) приводит к байесовскому правилу обучения:

$$(\hat{a}_1, \dots, \hat{a}_n, \hat{b}) = \arg \max [\ln \Psi(a_1, \dots, a_n | \mu) + \ln \Phi(X | Y, a_1, \dots, a_n, b)]. \quad (1.1)$$

Нетрудно показать, что при такой постановке задачи обучения (1.1) получим следующий оптимизационный критерий:

$$\begin{cases} -\ln \Psi(a_1, \dots, a_n | \mu) + c \sum_{j=1}^N \delta_j \rightarrow \min(a_1, \dots, a_n, b, \delta_1, \dots, \delta_N), \\ y_j (\sum_{i=1}^n a_i x_{ij} + b) \geq 1 - \delta_j, \delta_j \geq 0, j = 1, \dots, N. \end{cases} \quad (1.2)$$

В частности, если в качестве априорного распределения компонент направляющего вектора $\Psi(a_1, \dots, a_n | \mu) = \Psi(a_1, \dots, a_n)$ принять нормальное распределение с независимыми компонентами, нулевыми математическими ожиданиями и равными дисперсиями r , и обозначить $C = 2rc$, то получим классический *метод опорных векторов* (SVM) [1] для вещественных признаков $x_{ij} \in \mathbb{X}_i = \mathbb{R}$ и вещественных компонент направляющего вектора $a_i \in \mathbb{X}_i = \mathbb{R}$:

$$\begin{cases} \sum_{i=1}^n a_i^2 + C \sum_{j=1}^N \delta_j \rightarrow \min(a_1, \dots, a_n, b, \delta_1, \dots, \delta_N), \\ y_j (\sum_{i=1}^n a_i x_{ij} + b) \geq 1 - \delta_j, \delta_j \geq 0, j = 1, \dots, N. \end{cases} \quad (1.3)$$

В терминах потенциальных функций $K_i(x'_i, x''_i) : \mathbb{X}_i \times \mathbb{X}_i \rightarrow \mathbb{R}$, заданных в произвольных шкалах соответствующих модальностей $x_i \in \mathbb{X}_i$, метод опорных векторов (1.3) примет вид:

$$\begin{cases} \sum_{i=1}^n K_i(a_i, a_i) + C \sum_{j=1}^N \delta_j \rightarrow \min(a_1, \dots, a_n, b, \delta_1, \dots, \delta_N), \\ y_j (\sum_{i=1}^n K_i(a_i, x_{ij}) + b) \geq 1 - \delta_j, \delta_j \geq 0, j = 1, \dots, N. \end{cases} \quad (1.4)$$

Компоненты направляющего вектора a_i могут быть не представлены в исходной шкале значений признаков $a_i \notin \mathbb{X}_i$, а принадлежат гипотетическому линейному пространству $a_i \in \tilde{\mathbb{X}}_i \supseteq \mathbb{X}_i$ в которое исходная шкала значений погружена заданной потенциальной функцией. Однако всегда оказывается, что оптимальный направляющий элемент \hat{a}_i является линейной комбинацией реально существующих объектов из обучающей совокупности $\hat{a}_i = \sum_{j: \hat{\lambda}_j > 0} \hat{\lambda}_j y_j x_{ij} \in \tilde{\mathbb{X}}_i$, а решающее правило $\sum_{j: \hat{\lambda}_j > 0} \hat{\lambda}_j y_j \sum_{i=1}^n K_i(x_{ij}, x_i) + b \leq 0$ полностью определяется неотрицательными множителями Лагранжа $\hat{\lambda}_j > 0$ при ограничениях оптимизационной задачи (1.4), соответствующих *опорным объектам*.

В последующих двух главах рассматриваются два вида априорных распределений компонент направляющего вектора разделяющей гиперплоскости $\Psi(a_1, \dots, a_n | \mu)$, приводящих к двум модификациям метода опорных векторов, в которых параметр μ регулирует степень селективности отбора признаков.

2. МЕТОД РЕЛЕВАНТНЫХ ПОТЕНЦИАЛЬНЫХ ФУНКЦИЙ С РЕГУЛИРУЕМОЙ СЕЛЕКТИВНОСТЬЮ

Примем в качестве априорной плотности распределения компонент a_i направляющего вектора разделяющей гиперплоскости $y(x_1, \dots, x_n | a_1, \dots, a_n, b) \geq 0$ нормальное распределение с нулевыми математическими ожиданиями и дисперсиями r_i :

$$\begin{aligned} \psi(a_i | r_i) &= \left(1/r_i^{1/2} (2\pi)^{1/2}\right) \exp\left(-(1/2r_i) a_i^2\right), \\ \Psi(a_1, \dots, a_n | r_1, \dots, r_n) &\propto \left(\prod_{i=1}^n r_i\right)^{-1/2} \exp\left(-(1/2) \sum_{i=1}^n (1/r_i) a_i^2\right). \end{aligned}$$

Кроме того, будем предполагать, что и сами величины обратные дисперсиям $1/r_i$ имеют априорное гамма распределение

$$\gamma((1/r_i) | \alpha, \beta) \propto (1/r_i)^{\alpha-1} \exp(-\beta(1/r_i))$$

с равными математическими ожиданиями $E(1/r_i) = \alpha/\beta$ и дисперсиями $E((1/r_i)^2) = \alpha/\beta^2$, а также примем $\alpha = (1 + \mu)^2/2\mu$ и $\beta = 1/2\mu$. Таким образом получаем параметрическое семейство распределений относительно параметра $\mu \geq 0$, такое что $E(1/r_i) = (1 + \mu)^2$ и $E((1/r_i)^2) = 2\mu(1 + \mu)^2$. Если $\mu \rightarrow 0$, то значения $1/r_i$ уравниваются $1/r_1 \cong \dots \cong 1/r_n \cong 1$, однако, при увеличении μ величины $1/r_i$ могут значительно различаться, а при $\mu \rightarrow \infty \Rightarrow 1/r_i \rightarrow 0$.

Апостериорная совместная плотность распределения величин $1/r_i$ примет вид:

$$G(r_1, \dots, r_n | \mu) \propto \left(\prod_{i=1}^n 1/r_i \right)^{\alpha-1} \exp \left(-\beta \sum_{i=1}^n (1/r_i) \right)$$

Принцип максимизации совместной апостериорной плотности

$$P(a_1, \dots, a_n, b, r_1, \dots, r_n | X, Y, \mu) \propto \Psi(a_1, \dots, a_n | r_1, \dots, r_n) G(r_1, \dots, r_n | \mu) \Phi(X | Y, a_1, \dots, a_n, b)$$

приводит к критерию обучения:

$$\begin{cases} \sum_{i=1}^n [(1/r_i)(a_i^2 + (1/\mu)) + ((1/\mu) + 1 + \mu) \ln r_i] + \\ C \sum_{j=1}^N \delta_j \rightarrow \min(a_i, r_i, b, \delta_j), \\ y_j (\sum_{i=1}^n a_i x_{ij} + b) \leq 1 - \delta_j, \delta_j \geq 0, j = 1, \dots, N. \end{cases} \quad (2.1)$$

Небольшие значения r_i соответствуют небольшим по величине компонентам a_i и, следовательно, i -й признак фактически не участвует в решающем правиле $\sum_{i=1}^n a_i x_i + b \geq 0$, что легко видеть, если записать решающее правило через потенциальные функции $\sum_{j: \lambda_j > 0} \lambda_j y_j \sum_{i=1}^n r_i K_i(x_{ij}, x_i) + b \geq 0$, в котором дисперсии r_i выступают в роли весов соответствующих потенциальных функций.

Для решения оптимизационной задачи (2.1) предлагается проводить поочередную минимизацию по группам переменных (a_1, \dots, a_n, b) и (r_1, \dots, r_n) .

При $\mu = 0$ критерий (2.1) примет вид классического SVM (1.3) не обладающего селективностью, но при $\mu \rightarrow \infty$ получим критерий $\sum_{i=1}^n [(1/r_i)a_i^2 + \mu \ln r_i] + C \sum_{j=1}^N \delta_j \rightarrow \min$ в (2.1), который будет более селективным чем исходный критерий $\sum_{i=1}^n [(1/r_i)a_i^2 + \ln r_i] + C \sum_{j=1}^N \delta_j \rightarrow \min$ РКМ [8, 9], в следствии чего предлагаемый метод отбора признаков был назван *методом релевантных потенциальных функций с регулируемой селективностью*.

3. МЕТОД ОПОРНЫХ ПОТЕНЦИАЛЬНЫХ ФУНКЦИЙ С РЕГУЛИРУЕМОЙ СЕЛЕКТИВНОСТЬЮ

Теперь пусть априорная плотность $\Psi(a_1, \dots, a_n | \mu)$ выражена через выпуклую функцию $q(a | \mu)$ в виде:

$$\Psi(a_1, \dots, a_n | \mu) \propto \exp \left(- \sum_{i=1}^n q(a_i | \mu) \right).$$

Очевидно, что общий критерий обучения (1.2) примет вид:

$$\begin{cases} \sum_{i=1}^n q(a_i|\mu) + c \sum_{j=1}^N \delta_j \rightarrow \min(a_1, \dots, a_n, b, \delta_1, \dots, \delta_N), \\ y_j (\sum_{i=1}^n a_i x_{ij} + b) \leq 1 - \delta_j, \delta_j \geq 0, j = 1, \dots, N. \end{cases} \quad (3.1)$$

Для действительных признаков $x_i \in \mathbb{R}$ предлагается кусочно-линейная квадратичная функция:

$$q(a_i|\mu) = \begin{cases} 2\mu|a_i|, & \text{если } |a_i| \leq \mu, \\ a_i^2 + \mu^2, & |a_i| > \mu. \end{cases} \quad (3.2)$$

В терминах потенциальных функций эквивалентная запись функции имеет вид:

$$q(a_i|\mu) = \begin{cases} 2\mu \sqrt{K_i(a_i, a_i)}, & \text{если } \sqrt{K_i(a_i, a_i)} \leq \mu, \\ \mu^2 + K_i(a_i, a_i), & \text{если } \sqrt{K_i(a_i, a_i)} > \mu. \end{cases}$$

С учетом (3.2) оптимизационная задача (3.1) является задачей выпуклого программирования:

$$\begin{cases} 2\mu \sum_{|a_i| \leq \mu} |a_i| + \sum_{|a_i| > \mu} (a_i^2 + \mu^2) + c \sum_{j=1}^N \delta_j \rightarrow \min(a_1, \dots, a_n, b, \delta_1, \dots, \delta_N), \\ y_j (\sum_{i=1}^n a_i x_{ij} + b) \leq 1 - \delta_j, \delta_j \geq 0, j = 1, \dots, N. \end{cases} \quad (3.3)$$

Параметр $0 \leq \mu < \infty$ выполняет роль параметра селективности и при $\mu = 0 \Rightarrow q(a_i|\mu) = \text{const} + a_i^2$ оптимизационный критерий (3.3) эквивалентен SVM [1], а при $\mu \rightarrow \infty \Rightarrow q(a_i|\mu) \propto \mu|a_i|$ представляет собой метод SKM [9, 10] с увеличивающейся способностью отбора признаков по мере увеличения μ относительно параметра c .

Такой подход был назван *методом опорных потенциальных функций с регулируемой селективностью*, поскольку аналогично методу опорных потенциальных функций (SKM) в результате обучения при заданном значении параметра селективности $\mu \geq 0$ формируется подмножество $\hat{I}_\mu \subseteq I = \{1, \dots, n\}$ *опорных признаков (потенциальных функций)* с ненулевыми компонентами $\hat{a}_i^2 > 0$ направляющего вектора искомой разделяющей гиперплоскости.

4. ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ

В качестве модели двух классов объектов использовались два равномерных распределения в соприкасающихся гиперкубах $\mathbb{R}^n, n = 100$, расположенных по разные стороны относительно заданной гиперплоскости $\mathbf{a}^T \mathbf{x} \geq 0$ и ориентированных вдоль ее направляющего вектора $\mathbf{a} = (a_1 = 0.8, a_2 = 0.75, \dots, a_5 = 0.6, a_6 = 0, \dots, a_{100} = 0)$, в котором первые 5 элементов отличны от нуля. Таким образом только 5 признаков содержат информацию о разделении классов и 95 являются шумовыми. Генерировались обучающие выборки размером $N = N_{+1} = N_{-1} = 50 + 50 = 100$ и тестовые $N_{test} = 5000 + 5000 = 10000$.

Для набора возрастающих значений параметра селективности $\mu \geq 0$ проводилось обучение по критериям (2.1) и (3.3), и оценивалась обобщающая способность полученных решений на тестовой совокупности (ошибку на тестовой совокупности для больших N_{test} можно считать ошибкой на генеральной совокупности), а так же на скользящем контроле (10-fold validation). Результаты экспериментов приведены на рисунке 1.

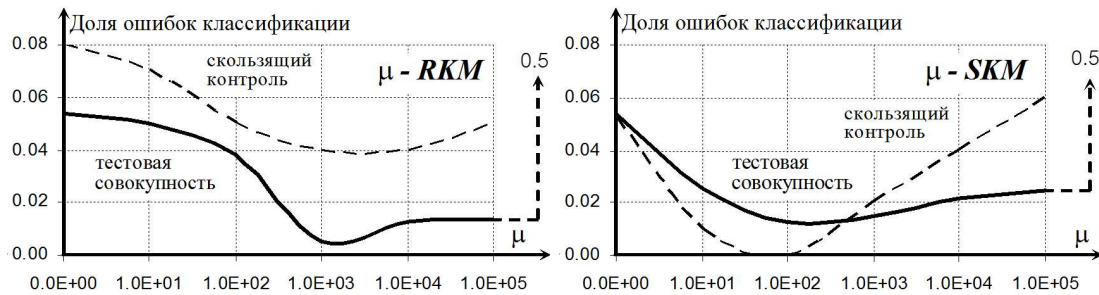


Рис. 1. Качество обучения на тестовой совокупности и по результатам скользящего контроля (10-fold validation) для возрастающего набора значений параметра селективности $0 \rightarrow \mu \rightarrow \infty$.

Ошибка на генеральной совокупности обычного SVM (1.3) на первых 5 "разумных" признаках составляет 0.0045, а только на первом самом информативном признаке a_1 в отдельности – 0.0245.

При наименьшей селективности $\mu = 0$ ошибка обоих методов на тестовой совокупности составляет 0.058, что эквивалентно обучению SVM на всех 100 признаках одновременно.

Минимальная ошибка, достигаемая методом RKM, составляет 0.0052, что сравнимо с ошибкой SVM 0.0045 на первых 5 "разумных" признаках. Однако ошибка метода SKM 0.0124 в два раза больше чем у RKM, что можно объяснить недостаточной гибкостью метода с отбором подмножества опорных признаков в сравнении с взвешиванием признаков.

При дальнейшем увеличении селективности $\mu \rightarrow \infty$ все веса у RKM стремятся к нулю $r_i \rightarrow 0$, а SKM сокращает множество опорных признаков до пустого множества. Соответственно, ошибка обучения сначала достигает критического уровня 0.0245, а затем ошибка поднимается до максимального уровня 0.5.

Таким образом, полученные результаты наглядно демонстрируют адекватность механизма регулируемой селективности для повышения обобщающей способности при многомодальном распознавании образов.

Работа выполнена при поддержке РФФИ, проекты № 05-01-00679, 06-01-08042, 06-07-89249.

СПИСОК ЛИТЕРАТУРЫ

1. Ross A., Jain A. Multimodal biometrics: An overview. // Proceedings of the 12th European Signal Processing Conference, Vienna, Austria, 2004. – С. 1221.
2. Jannin P, Fleig O.J, Seigneuret E, Grova C, Morandi X, Scarabin J.M. A data fusion environment for multimodal and multi-informational neuronavigation. // Computer Aided Surgery, 2000, Vol. 5, No. 1, pp. 1-10.
3. M.A. Aizerman, E.M. Braverman, L.I. Rozonoer. Theoretical foundations of the potential function method in pattern recognition learning. // Automation and Remote Control, 1964, Vol. 25, pp. 821-837.
4. Vapnik V. Statistical Learning Theory. John-Wiley // & Sons, Inc. 1998.

5. *Duin, R.P. W, De Ridder, D., Tax, D.M.J.* Featureless classification. // Proceedings of the Workshop on Statistical Pattern Recognition, Prague, June 1997.
6. *V. Mottl.* Metric spaces admitting linear operations and inner product. // Doklady Mathematics 67(1), 2003, 140–143.
7. *Guyon I. M., Gunn S. R., Nikravesh M., Zadeh L., Eds.* Feature Extraction, Foundations and Applications. // Springer, 2006.
8. *Sulimova V., Mottl V., Tatarchuk A.* Multi-kernel approach to on-line signature verification. // Proceedings of the 8th IASTED International Conference on Signal and Image Processing. Honolulu, Hawaii, USA, August 14-16, 2006.
9. *Mottl V., Tatarchuk A., Sulimova V., Krasotkina O., Seredin O.* Combining pattern recognition modalities at the sensor level via kernel fusion. // Proceedings of the 7th International Workshop on Multiple Classifier Systems. Czech Academy of Sciences, Prague, Czech Republic, May 23-25, 2007.
10. *Sonnenburg S., Rätsch G., Schäfer C.* A general and efficient multiple kernel learning algorithm. // Proceedings of the 19th Annual Conference on Neural Information Processing Systems, Vancouver, Canada, December 5-8, 2005.
11. *Де Гроот М.* Оптимальные статистические решения. // Москва: Мир, 1974.

Статья поступила в редакцию 25.04.2008