ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2005, том 45, № 7, с. 1318–1325

УДК 519.712.63

ВЫПУКЛАЯ КЛАСТЕРНАЯ СТАБИЛИЗАЦИЯ АЛГОРИТМОВ РАСПОЗНАВАНИЯ КАК СПОСОБ ПОЛУЧЕНИЯ КОЛЛЕКТИВНЫХ РЕШЕНИЙ С ВЫСОКОЙ ОБОБЩАЮЩЕЙ СПОСОБНОСТЬЮ[1)]

© 2005 г. Д. П. Ветров, Д. А. Кропотов

(119991 Москва, ул. Вавилова, 40, ВЦ РАН)

e-mail: vetrovd@yandex.ru, dkropotov@yandex.ru

Поступила в редакцию 13.09.2004 г.

Обобщается метод получения коллективных решений в задачах распознавания образов, основанный на одновременном увеличении устойчивости и эффективности (т.е. процента правильно распознанных объектов обучающей выборки) работы. Показана связь между описанной процедурой и несколькими известными методами построения коллективов алгоритмов, представляющих собой частные случаи более общего подхода. Практическая полезность метода подтверждена на примере нескольких широко известных задач распознавания. Библ. 15. Табл. 1.

Ключевые слова: распознавание образов, коллективные решения, коррекция алгоритмов, устойчивость классификаторов.

1. ВВЕДЕНИЕ

Известно большое количество различных параметрических семейств алгоритмов распознавания образов. При решении конкретной практической задачи обычно неясно, какой метод распознавания наиболее полно учитывает ее особенности. Часто учет результатов распознавания несколькими методами позволяет улучшить качество классификации. В связи с этим большое распространение получили методы построения коллективных решений над разными семействами (см. [1], [2]), а также коррекция алгоритмов (см. [3], [4]). Следует отметить, что методы коррекции применимы для синтеза коллективных решений в рамках одного (быть может, достаточно широкого) алгоритмического семейства.

Использование представителей разных семейств классификаторов при построении коллективного решения, вообще говоря, позволяет преодолеть ограничения, свойственные той или иной параметрической модели. Тем не менее при этом возникают серьезные проблемы по унификации выходов алгоритмов различных семейств для их последующего объединения. Кроме того, существует реальная опасность дополнительной перенастройки получившегося коллективного решения, в особенности если оно имеет вид сложной зависимости от исходных алгоритмов. Это свидетельствует о необходимости учета дополнительных факторов помимо эффективности при синтезе коллективного решения. Ведь конечной целью при решении задачи распознавания образов является снижение ошибки обобщения, т.е. вероятности неправильной классификации произвольного объекта, взятого из генеральной совокупности. За последние годы достигнуты определенные успехи в оценивании взаимосвязи между эффективностью и ошибкой обобщения, опирающиеся на понятие емкости классификатора (ВЧ-размерность, см. [5]). Несмотря на фундаментальный характер разработанной с его помощью теории статистического обучения, оценкам, основанным на емкости классификатора, свойственен ряд недостатков (в частности, завышенный объем требуемых обучающих выборок, пессимистические оценки уровня ошибки обобщения, а также трудности при определении емкостей сложных классификаторов). Более того, при использовании коллективов алгоритмов из различных параметрических семейств подсчет емкости получившегося коллективного решения вообще не представляется возможным. Следовательно, необходимы другие характеристики, которые позволили хотя бы косвенно оценивать способность к обобщению как отдельных алгоритмов, так и их объединений. К таковым можно

1) Работа выполнена при частичной финансовой поддержке РФФИ (коды проектов 02-01-08007, 02-07-90134, 02-07-90137, 03-01-00580), целевой программы ОМН РАН № 02, гранта президента РФ: НШ-1721.2003.1 и госконтракта № 10002-251/П-17/026-024/200404-042.

отнести, например, требование устойчивости алгоритма, а именно: при небольших изменениях значений признаков объектов, результаты распознавания также должны слабо меняться. Заметим, что это не единственно возможное ограничение по соображениям устойчивости. Так, при байесовской регуляризации машинного обучения одним из ограничений является слабое изменение результатов распознавания при небольших изменениях параметров алгоритма (см. [6], [7]). В случае построения коллективных решений такой подход представляется весьма затруднительным, так как число параметров решающего правила становится весьма большим (к ним относятся параметры каждого алгоритма, входящего в коллектив, а также параметры самого коллективного решения), а кроме того, степень их влияния, вообще говоря, различна в разных подобластях пространства признаков. Заметим, что условие устойчивости неизбежно приводит нас к необходимости рассмотрения оценок за классы в качестве выхода алгоритма распознавания. Впрочем, как было показано в [3], требование наличия промежуточных оценок за классы, необходимое и для коррекции алгоритмов распознавания, не является слишком обременительным, поскольку любой метод классификации можно представить в виде композиции распознающего оператора, обладающего требуемыми свойствами, и решающего правила, возвращающего индекс класса, к которому данный объект был отнесен алгоритмом.

В работе [8] была предложена схема выпуклой стабилизации алгоритмов, входящих в коллектив. Были получены теоретические результаты о возможности построения коллективного решения, обладающего свойствами корректного алгоритма. Основные положения выпуклой стабилизации изложены ниже. В разд. 3 предлагается схема выпуклой кластерной стабилизации, являющаяся обобщением результатов выпуклой стабилизации. В разд. 4 приводятся результаты сравнительных экспериментов и их обсуждение.

2. ВЫПУКЛЫЙ СТАБИЛИЗАТОР

Будем рассматривать следующую задачу распознавания образов. Пусть имеется выборка из m объектов, каждый из которых принадлежит к одному из l непересекающихся классов. Каждый объект $\{x_i\}_{i=1}^m$ является вектором в n-мерном пространстве признаков. Задача состоит в отнесении нового объекта к одному из l классов так, чтобы его классификация соответствовала в наибольшей степени имеющейся выборке данных. Будем считать, что все рассматриваемые ниже алгоритмы распознавания представимы в виде

$$A(x) = \{P(\omega_j|x)\}_{j=1}^l = \mathbf{P}(\boldsymbol{\omega}|x). \qquad (1)$$

Другими словами, выходами алгоритма распознавания служат апостериорные вероятности принадлежности объъекта к каждому классу. Заметим, что не все алгоритмические модели оперируют в вероятностных терминах. Тем не менее, выходы большинства алгоритмов можно адекватно свести к вероятностным (например, для алгоритмов, основанных на построении гиперплоскости в пространстве признаков, достаточно взять логистическую функцию от расстояния между объектом и разделяющей гиперповерхностью). В дальнейшем будем полагать, что признаки соответствующим образом шкалированы (например, имеют единичную дисперсию).

Определение 1. Неустойчивостью алгоритма распознавания образов A в точке x называется величина

$$Z_A(x, \varepsilon) = \sum_{j=1}^{l}\left[P\left(\omega_j \mid x + \varepsilon\sum_{i=1}^{n}\mathbf{e}_i\right) - P(\omega_j \mid x)\right]^2,$$

где e_i – единичный вектор соответствующей координаты признакового пространства.

Таким образом, неустойчивость характеризует изменчивость результатов классификации при изменении координат распознаваемого объекта и при достаточно малых ε приближенно равна

$$Z_A(x, \varepsilon) \approx \varepsilon^2\|\nabla\mathbf{P}(\boldsymbol{\omega} \mid x)\|^2.$$

Предположим, что имеется p классификаторов $A_1, \ldots, A_p$, быть может принадлежащие разным алгоритмическим семействам, возвращающие вектора апостериорных вероятностей за классы. Для построения коллективного решения будем использовать контрольную (или валидационную) выборку $\{y_k\}_{k=1}^q$, которая, в частности, может совпадать с обучающей. Заметим, что использование дополнительной выборки для улучшения качества работы уже обученных алгоритмов является широкоиспользуемым приемом (например, так называемое pruning set, использующееся при пост-обработке решающих деревьев, см. [9] и др.). Обозначим

$$\Theta(k) = \{t \mid A_t \text{ правильно классифицирует } y_k\},$$

$$R(k) = \arg\min_{t \in \Theta(k)} Z_{A_t}(y_k), \quad T(k) = \arg\min_{t \in \{1, 2, \dots, p\}} Z_{A_t}(y_k),$$

$$P(k) = \begin{cases} R(k) & \text{при } \Theta(k) \neq \emptyset, \\ T(k) & \text{иначе.} \end{cases}$$

Определение 2. Алгоритм A получен из алгоритмов $A_1, \dots, A_p$ путем применения *выпуклого стабилизатора*, если его можно представить в виде следующей выпуклой комбинации исходных алгоритмов

$$P_A(\omega_j \mid x) = \frac{\sum_{k=1}^{q} w_k(x) P_{A_{F(k)}}(\omega_j \mid x)}{\sum_{k=1}^{q} w_k(x)}, \tag{2}$$

где $F : \{1, 2, \dots, q\} \longrightarrow \{1, 2, \dots, p\}$ – некоторая функция, определяющая индекс “наилучшего” алгоритма распознавания для каждого объекта контрольной выборки, а $w_k : \mathbb{R}^n \longrightarrow \mathbb{R}$ – весовые функции, обладающие следующими свойствами:

$$\begin{gathered} w_k(x) \geq 0 \quad \forall k = 1, 2, \dots, q, \\ w_k(x) \longrightarrow 0 \text{ при } \rho(x, y_k) \longrightarrow \infty, \\ \frac{w_k(x)}{\sum_{j=1}^{q} w_k(x)} \longrightarrow 1 \text{ при } \rho(x, y_k) \longrightarrow 0. \end{gathered} \tag{3}$$

В [8] было предложено брать $F(k) = P(k)$, а в качестве весовых функций использовать следующие выражения:

$$w_k(x) = \begin{cases} 1, & \rho(x, y_k) \leq \varepsilon, \\ 0, & \exists y_i \neq y_k : \rho(x, y_k) \leq \varepsilon, \\ \dfrac{1}{\rho(x, y_k) - \varepsilon}, & \rho(x, y_k) > \varepsilon \ \ \forall y_i. \end{cases}$$

Выпуклая стабилизация оказалась весьма эффективной при решении задач с малыми и сверхмалыми выборками (см. [8]). Однако использование описанного стабилизатора для задач с объемами выборок порядка сотни и выше сопряжено со значительными вычислительными трудностями. Кроме того, наличие большого количества объектов в контрольной выборке приводит к излишним “переключениям” между алгоритмами, что, в свою очередь, делает коллективное решение нестабильным, несмотря на предпринятые усилия.

3. ВЫПУКЛАЯ КЛАСТЕРНАЯ СТАБИЛИЗАЦИЯ

В основе выпуклой стабилизации лежат две идеи. Во-первых, коллективное решение ищется в виде выпуклой комбинации “лучших” в некоторой области признакового пространства классификаторов, веса которых зависят от положения распознаваемого объекта. Чем ближе он находится к той или иной области, тем выше вес соответствующего алгоритма в выпуклой комбинации. Фактически, это напоминает использование областей компетенции с нечеткими границами. Во-вторых, для каждой области наилучший алгоритм определяется, вообще говоря, не только показателем эффективности, но и дополнительным требованием устойчивости по отношению к небольшим вариациям координат объектов. В следующем разделе будет показано, что именно одновременное использование этих двух идей приводит к улучшению обобщающей способности построенного алгоритма. Построим обобщение выпуклого стабилизатора с учетом сформулированных выше концепций.

Разобьем контрольную выборку на r кластеров $D_1, \ldots, D_r$ и для каждого кластера введем понятие наилучшего алгоритма.

Определение 3. Будем говорить, что алгоритм A_{t_0} – *наилучший для k-го кластера*, если он доставляет максимум следующей целевой функции:

$$t_0 = \arg \max_{t = 1, 2, \ldots, p} [E_k(A_t) + \alpha S_k(A_t)],$$

где $E_k(A_t)$ – доля правильно классифицированных объектов контрольной выборки, принадлежащих k-му кластеру, алгоритмом A_t, а $S_k(A_t)$ – компонента устойчивости, определяемая следующим образом:

$$S_k(A) = \frac{1}{q_k} \sum_{y_i \in D_k} \exp\left(-\frac{Z_A(y_i, \varepsilon)}{2\lambda^2}\right).$$

В последней формуле q_k – количество объектов контрольной выборки, попавших в k-й кластер, λ и α – действительные числа такие, что $\lambda \geq 0$, а $\alpha > 0$.

Заметим, что такое определение наилучшего алгоритма для области пространства, образованной соответствующим кластером, представляет собой компромисс между требованием хорошего качества работы на объектах с известным ответом и требованием устойчивости, позволяющим надеяться на “перенесение” этого качества на произвольные объекты генеральной совокупности. Компромисс определяется параметрами α, ε и λ. Как будет указано ниже, каждый из параметров может быть определен либо в процессе обучения, либо его наилучшее значение (т.е. значение, при котором достигается наилучшее качество работы коллективного решения) практически не зависит от вида конкретной задачи.

Определение 4. Алгоритм A получен из алгоритмов $A_1, \ldots, A_p$ путем применения *выпуклого кластерного стабилизатора*, если его можно представить в виде следующей выпуклой комбинации исходных алгоритмов:

$$P_A(\omega_j \mid x) = \frac{\sum_{k=1}^{r} w_k(x) P_{A_{G(k)}}(\omega_j \mid x)}{\sum_{k=1}^{r} w_k(x)}, \tag{4}$$

где $G : \{1, 2, \ldots, r\} \longrightarrow \{1, 2, \ldots, p\}$ – функция, возвращающая индекс “наилучшего” алгоритма распознавания для соответствующего кластера, а $w_k : \mathbb{R}^n \longrightarrow \mathbb{R}$ – весовые функции, обладающие следующими свойствами:

$$\begin{gathered} w_k(x) \geq 0 \quad \forall k = 1, 2, \ldots, r, \\ w_k(x) \longrightarrow 0 \text{ при } \rho(x, C_k) \longrightarrow \infty, \\ \frac{w_k(x)}{\sum_{j=1}^{q} w_k(x)} \longrightarrow 1 \text{ при } \rho(x, C_k) \longrightarrow 0. \end{gathered} \tag{5}$$

В последних формулах $C_k = \frac{1}{q_k}\sum_{y_i \in D_k} y_i$ – центры кластеров.

Простейшим примером весовых функций являются следующие:

$$w_k(x) = \begin{cases} 1 \text{ при } \rho(x, C_k) < \rho(x, C_i) \quad \forall i \neq k, \\ 0 \text{ иначе.} \end{cases} \tag{6}$$

Легко видеть, что такая система весовых функций задает четкое разбиение пространства на непересекающиеся области. Она может быть полезна при решении существенно дискретных задач (т.е. задач с номинальными признаками или малым числом дискретных состояний).

Покажем теперь, что предлагаемая схема является обобщением некоторых известных методов построения коллективного решения.

Теорема 1. *Коллективное решение, полученное в результате применения выпуклого кластерного стабилизатора* (4), (6) *при* $\alpha = 0$, *совпадает с результатом применения к исходному коллективу метода областей компетенции* (см. [10], [11]).

Доказательство. Обозначим через A результат применения выпуклого кластерного стабилизатора с указанными параметрами к коллективу алгоритмов $A_1, \ldots, A_p$. В силу (6) очевидно, что для любого объекта результат применения к нему такого стабилизатора совпадает с результатом работы одного из членов коллектива, признанного наилучшим для данной области ($A(x) = A_{t_0}(x)$). При $\alpha = 0$ компонента устойчивости не учитывается при определении наилучшего алгоритма. Таким образом, наилучшим алгоритмом для заданной области является тот член коллектива, который допускает меньше всего ошибок на объектах контрольнгой выборки, принадлежащих данному кластеру. Но такой способ получения коллективного решения совпадает с методом областей компетенции, известным как Clustering & Selection, описанным, например, в [11]. Теорема полностью доказана.

Теорема 2. *Коллективное решение, полученное в результате применения выпуклого кластерного стабилизатора* (4), (5) *при* $r = q$ и $0 < \alpha < 1$, *совпадает с результатом применения к коллективу выпуклого стабилизатор*а (2), (3) *с той же системой весовых функций и* $F(k) = P(k)$.

Доказательство. Очевидно, что при числе кластеров, равных количеству объектов контрольной выборки, в каждый кластер попадет ровно один объект, который и будет являться центром кластера. Следовательно, условия (3) эквивалентно условиям (5). Отсюда вытекает возможность использования идентичных весовых функций в обоих случаях, т.е. каждой системе весовых функций выпуклого стабилизатора можно сопоставить эквивалентную ей систему функций выпуклого кластерного стабилизатора и наоборот.

Осталось показать, что функция $G(k) = P(k)$. Пусть для объекта $y_k = C_k$ найдется бы один алгоритм в коллективе, который распознает его правильно. Тогда $P(k)$ будет равна индексу наиболее устойчивого из них, т.е. обладающего наименьшим значением $Z(y_k, \varepsilon)$. Рассмотрим теперь выражение $V(A_t) = E_k(A_t) + \alpha S_k(A_t)$. Для всех алгоритмов, неправильно классифицирующих y_k, значение $V(A_t) < 1$, так как $E_k(A_t) = 0$, а $S_k(A_t) = \exp\left(-\dfrac{Z_{A_t}(y_k, \varepsilon)}{2\lambda^2}\right) \in (0, 1]$ и $0 < \alpha < 1$. Любой алгоритм, правильно классифицирующий объект y_k, имеет $V(A_t) > 1$, так как $E_k(A_t) = 1$. Следовательно поиск наилучшего в k-м кластере осуществляется среди алгоритмов, чьи индексы входят в множество $\Theta(k)$. Наилучшим алгоритмом среди них является классификатор, обладающий наибольшим значением $S_k(A)$. Так как значение $S_k(A)$ монотонно убывает по $Z_A(y_k, \varepsilon)$ при любых λ, то наибольшее значение соответствует наименьшему значению неустойчивости на объекте y_k. Следовательно, в случае если $\Theta(k) \neq \emptyset$, то $G(k) = R(k) = P(k)$.

В случае когда ни один алгоритм коллектива не может правильно классифицировать объект, аналогичными рассуждениями, учитывая что $E_k(A_t) = 0 \ \forall t$, можно показать, что $G(k) = T(k) = P(k)$. Теорема полностью доказана.

Последняя теорема означает, что выпуклый стабилизатор можно определить как частный случай выпуклого кластерного стабилизатора.

4. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

Для проверки практической пригодности изложенного метода была проведена серия экспериментов по построению коллективных решений. Для тестирования использовались широкоизвестные задачи распознавания, взятые из хранилища UCI repository (см. [12]). Следует отметить, что эти задачи часто используются для сравнения различных парадигм распознавания, представляя собой де-факто полигон для испытания различных идей. Эксперименты проводились с помощью программного комплекса “РАСПОЗНАВАНИЕ” (см. [13]) – универсального программного средства для исследования различных задач и методов анализа данных. Во всех случаях использовался коллектив, состоящий из шести методов: Линейный дискриминант Фишера, однослойный персептрон с 10 нейронами в слое, метод q-ближайших соседей и три метода опорных векторов с различными параметрами (линейная разделяющая гиперплоскость, кубическая гиперповерхность и случай гауссовских ядровых функций). Выходы всех методов были модифицированы с учетом требования (1). В качестве альтернативных методов построения коллективных решений использовались хорошо зарекомендовавшие себя алгоритмы, а именно метод Байеса (см. [14]) и метод шаблонов принятия решений с Евклидовой метрикой (см. [15]). Остальные коллективные методы, заложенные в систему “РАСПОЗНАВАНИЕ” (комитетные методы, метод Вудса, четкие области компетенции и др.), уступают названным выше на данных задачах.

Во всех задачах в качестве контрольной выборки использовалась обучающая. Весовые функции определялись соотношениями

$$w_k(x) = \frac{1}{\rho^2(x, C_k)}\exp\left(-\frac{\rho^2(x, C_k)}{2\sigma^2}\right). \tag{7}$$

Параметры масштабировки были равны характерному расстоянию в выборке

$$\lambda = \sigma = \varepsilon = \max\min_{i \neq j}\rho(y_i, y_j),$$

а количество кластеров r и значение α подбирались с помощью независимой выборки либо по скользящему контролю.

Результаты экспериментов показаны в таблице. В колонке SB указан результат работы (процент правильно распознанных объектов) наилучшего метода коллектива. Далее указано изменение результата после применения метода Байеса NB, шаблонов принятия решений DT, выпуклого кластерного стабилизатора CCS. В ходе экспериментов исследовалось также влияние нечеткости подобластей признакового пространства и принципа устойчивости на обобщающую способность. В колонке CA указано изменение результата после применения классических областей компетенции (т.е. выпуклого кластерного стабилизатора с $\alpha = 0$ и весовыми функциями, определяемыми выражением (6)). В колонке CC указаны изменения результата после применения выпуклого кластерного стабилизатора с весовыми функциями (7) и $\alpha = 0$. Наконец, в колонке SA указано изменение после применения стабилизатора с четкими границами областей (т.е. весовыми функциями (6) и значением α, равным соответствующей величине в методе CCS). В таблице в каждой колонке, соответствующей результатам применения коллективных решений, жирным шрифтом выделены случаи, когда коллективное решение приводит к улучшению ситуации.

Таблица

Задача	SB	NB	DT	CCS	CA	CC	SA
Melanoma	59.4	3.1	–3.1	6.2	0	3.1	–6.3
Breast	95.5	–1.4	0	–1.1	–1.1	–1.1	–1.1
Credit	86.2	–5.5	–3.7	0.3	–10.9	–1.1	–5.9
Eco	82.7	–2.8	1.1	0.5	–3.9	–0.6	–11.2
Hea	57.4	–5.9	–3	0	–0.1	–0.5	–6.7
Image	86.7	4.4	3.1	4.2	–0.2	1.1	–13.2
Yea	59.7	–1.3	–0.5	–0.1	–5.8	–0.1	–2.8

Результаты экспериментов позволяют сделать несколько выводов. Во-первых, очевидным является значительное увеличение качества распознавания при использовании нечеткого разбиения пространства признаков. Во-вторых, в большинстве задач учет условия стабильности также улучшает качество распознавания. Более того, одновременное использование нечетких областей и условия устойчивости еще увеличивает качество работы. Можно отметить, что выпуклый кластерный стабилизатор во всех случаях показал высокое качество работы, часто превосходящее результаты лучших коллективных методов. Объем требуемых вычислений значительно меньше, чем при использовании обычного выпуклого стабилизатора. В-третьих, существует ряд задач (например, Breast), на которых ни один из использовавшихся коллективных методов не смог улучшить результат, полученный с помощью лучшего члена коллектива. Более того, использование коллективных методов только ухудшает ситуацию. Вероятно, это связано с тем, что в данных задачах наилучший метод учитывает все значимые закономерности, содержащиеся в данных, и учет результатов других методов не добавляет никакой полезной информации, делая при этом модель более сложной, а значит, и более подверженной случайным возмущениям. Также можно отметить еще одну любопытную деталь. В ходе экспериментов оказалось, что наилучшее значение числа кластеров приблизительно пропорционально квадратному корню из объема контрольной выборки

$$r \sim \beta\sqrt{q}$$

с $\beta \in (0.5, 3)$. По-видимому, можно провести аналогию между этим фактом и задачей восстановления вероятностных плотностей в непараметрической статистике методом динамических окон, в которых число объектов в окне также рекомендуется брать пропорциональным квадратному корню из объема выборки.

5. ЗАКЛЮЧЕНИЕ

В настоящей статье была предложена обобщенная схема построения коллективных решений над множеством алгоритмов, принадлежащих к разным семействам. Основной целью при ее разработке было увеличение обобщающей способности получившегося алгоритма без привлечения дополнительной выборки прецедентов. Для этого была использована поправка к функционалу, определяющему степень пригодности алгоритма в данной подобласти, отвечающая за устойчивость результата при изменениях координат объекта. При построении коллективного решения использовалась парадигма областей компетенции, которая была обобщена на случай нечетких границ между областями. Многочисленные эксперименты показали, что учет устойчивости метода в области и использование нечетких переходов между областями, задаваемых системой весовых функций, позволяют добиться улучшения качества работы получившегося коллективного решения как по сравнению с исходными алгоритмами, так и по сравнению с другими гибридными схемами.

СПИСОК ЛИТЕРАТУРЫ

1. *Kittler J., Hatef M., Duin R., Matas J.* On combining classifiers // IEEE Trans. Pattern Analys. and Mach. Intelligence. 1998. V. 20. № 3. P. 226–239.
2. *Вешторт А.М., Зуев Ю.А., Краснопрошин В.В.* Двухуровневая схема распознавания с логическим корректором // Распознавание, классификация, прогноз. Матем. методы и их применение. 1989. Вып. 2. С. 73–98.
3. *Журавлев Ю.И.* Избранные научные труды. М.: Магистр, 1998.
4. *Shapire R., Freund Y., Bartlett P., Lee W.* Boosting the margin: a new explanation for the effectiveness of voting methods // Proc. 14th Internat. Conf. Mach. Learning. P. 322–330.
5. *Вапник В.Н., Червоненкис А.Я.* Теория распознавания образов. М.: Наука, 1974.
6. *Шумский С.А.* Байесовская регуляризация обучения // IV Всерос. научно-техн. конф. “Нейроинформатика 2002”. 2002. Т. 2. С. 30–93.
7. *MacKay D.* Bayesian interpolation // Neural Comput. 1992. V. 4. № 3. P. 415–447.
8. *Ветров Д.П.* О синтезе корректных алгоритмов распознавания образов с минимальной величиной неустойчивости // Ж. вычисл. матем. и матем. физ. 2003. Т. 3. № 11.С. 1754–1760.
9. *Esposito F., Malerba D., Semeraro G.* A compararive analysis of methods for pruning decision trees // IEEE Trans. Pattern Analys. Mach. Intelligence. 1997. V. 19. № 5. P. 476–492.

10. *Расстригин Л.А., Эренштейн Р.Х.* Метод коллективного распознавания. М.: Энергоиздат, 1981.
11. *Kopustinskas A., Lipnikas A.* Classifier fusion with data-dependent aggregation schemes. Proc. 7th Internat. Conf. Inform. Networks, Systems and Technol. 2001. V. 1. P. 147–153.
12. *Murphy P., Aha D.* UCI repository of machine learning databases // Univ. California, Dept. Informat. and Comput. Sci. California: Irvine, 1996.
13. *Журавлев Ю.И., Рязанов В.В., Сенько О.В., и др.* Разработка универсальной программной системы интеллектуального анализа данных, распознавания и прогноза // Докл. XI Всерос. конф. Матем. методы распознавания образов. 2003. С. 311–314.
14. *Xu L., Krzyzak A., Suen C.* Methods of combining multiple classifiers and their application to handwritten recognition // IEEE Trans. Systems, Man, Cybernetics. 1992. V. 22. P. 418–435
15. *Kuncheva L., Bezdek J., Duin R.* Decision templates for multiple classifier fusion: an experimental comparison // Pattern Recognition. 2001. V. 34. № 2. P. 299–314.