

УДК 004.045

**Построение субоптимального метода
распознавания и классификации
на смесевых моделях вероятностных распределений
для обработки дистанционных данных**

А. И. Боровик, С. В. Деркач, И. А. Колтунов

Харьковский национальный университет имени В.Н. Каразина, Украина

На смесевых моделях вероятностных распределений для признаков распознаваемых объектов строится субоптимальный метод *распознавания и классификации* (РИК). Оптимизация проводится на двух многокомпонентных факторах: на множестве признаков-характеристик объектов и на множестве правил классификации. Для обеспечения непрерывности траектории оптимизации метода РИК вводится понятие топологической окрестности точки на гиперплоскости с гиперосями: «характеристики» и «правила».

Ключевые слова: метод распознавания и классификации, вероятностные модели распределения, нижние доверительные границы принадлежности классам.

На сумішевих моделях імовірнісних розподілів для ознак розпізнаваних об'єктів будується субоптимальних метод розпізнавання і класифікації (РІК). Оптимізація проводиться на двох багатоконпонентних факторах: на множині ознак-характеристик об'єктів і на множині правил класифікації. Для забезпечення безперервності траєкторії оптимізації методу РІК вводится поняття топологічного околу точки на гіперплоскості з гіперосями: «характеристики» та «правила».

Ключові слова: метод розпізнавання та класифікації, імовірнісні моделі розподілу, нижні довірчі межі приналежності класам.

On composite models of probability distributions for characteristics of recognizable objects suboptimal recognition and classification method (RAC) is constructed. Optimization is carried out on two multi-factors: on the set of attributes, characteristics of objects and on the set of classification rules. To ensure the continuity of the trajectory optimization of the method RAC the concept of topological neighborhood of the hyper plane with hyper axis is introduced: "characteristics" and "rules".

Key words: recognition and classification method, the probability distribution model, the lower confidence limits membership in the class.

1. Общая постановка задачи и её актуальность

Компьютерные технологии «методы РИК» являются составной частью актуальной научно-технической проблемы современности – создания «искусственного интеллекта». В настоящее время невозможно назвать область хозяйственных, научно-технических, учебно-образовательных, гуманитарных и культурных интересов человека, где не используются «методы РИК» [1,2]. Особую практическую значимость приобретают методы РИК в высоких технологиях: электроника, робототехника, видео- и радиотехника, био- и нанотехнологии, дешифрирование сигналов, полей и изображений, дистанционные космические исследования.

Сложность математической проблемы РИК заключается: а) в отсутствии общей математической модели выбора правила классификации объектов и выбора распознающих признаков объектов; б) в неустойчивости большинства частных математических моделей РИК, представляющих собой класс так называемых некорректно поставленных обратных задач; в) в многомерности пространства распознающих признаков для классифицируемых объектов [3,4,5]. По сравнению с десятками известных методов РИК несравненно лучшими возможностями обладают статистические методы РИК[3-7].

Метод РИК называют статистическим, если при алгоритмическом обучении (частичном обучении, самообучении) определяются *законы распределения вероятностей* (ЗРВ) распознающих признаков для объектов изучаемых классов, а при алгоритмическом распознавании оцениваются вероятности принадлежности предъявляемых объектов этим классам [3,7]. Перечислим некоторые положительные особенности статистических методов РИК[7].

В большинстве приложений экспериментальные данные являются флуктуативными, притом настолько сильно, что области пересечения различных классов в пространстве признаков существенно пересекаются, вплоть до полного вложения одних классов в другие. Статистические методы позволяют решать вполне удовлетворительно даже такие задачи.

Классифицируемые объекты оказываются существенно неоднородными в каждом из заданных классов. В статистических моделях распознавания это означает, например, многомодальность распределений признаков объектов в каждом классе; в детерминированном описании это означает многосвязность областей задания класса в признаковом пространстве. Достаточно развитые статистические модели (например, СМВР) дают решение этой проблеме.

При формировании ОВ нет полного (на 100%) знания о принадлежности объектов заданным классам. Количественную меру субъективной неуверенности дают те же статистические методы.

Объект может объективно принадлежать различным классам, как физическая смесь. Нечеткая принадлежность объекта тоже может быть описана СМВР, как на уровне постановки задачи, так и при интерпретации результатов.

Принадлежность объекта различным классам может оказаться логически непротиворечивым утверждением, однако неизвестным исследователю. Установлению этого факта могут способствовать статистические методы со смесевыми моделями.

Адекватность достаточно сложного статистического описания реальному распределению объектов каждого класса приводит к ненужности формирования так называемой контрольной выборки, потому что ККК на обучающей выборке не имеет «оптимистического смещения».

На несмещенном ККК можно одновременно решить обе задачи оптимизации: построение оптимального набора распознающих признаков и нахождение оптимального алгоритма классификации в семействе заданных статистических алгоритмов. Важно отметить, что излишняя структуризация алгоритма классификации, в том числе использование ненужных признаков - это не только технологическое неудобство. Главное: при ограниченном объеме ОВ - а это всегда так - ухудшается качество распознавания.

Замкнутость относительно линейных преобразований широкого класса статистических методов, основанных на полигауссовых распределениях [7], позволяет решить многие проблемы выбора распознающих признаков: переход к субоптимальным линейным комбинациям признаков, сжатие исходных данных использованием спектральных признаков, учет текстуры сигналов и полей формированием текстурных признаков.

Та же замкнутость относительно линейных преобразований дает возможность производить классификацию объектов с дискретно распределенными и качественными признаками путем погружения дискретных распределений в полигауссову смесь случайной раскачкой обучающей выборки.

Использование достаточно сложного «адекватного» реальной ситуации алгоритма классификации позволяет получить достоверную оценку классификации для каждого объекта в отдельности, а не среднюю достоверность для ансамбля объектов. При высокой цене ошибки классификации отдельных объектов это необходимое условие практической применимости метода.

Применение статистических алгоритмов распознавания для обработки именно изображений имеет свои особенности: например, когда объектами для обучения являются пиксели изображения, то их количество обычно достаточно для репрезентативности обучающей выборки.

Применение сложных статистических распределений приводит к вынужденному отказу от заведомо упрощенных аналитических построений. Многие вопросы предварительного исследования, в том числе в выборе оптимальных параметров алгоритмов, решаются численными методами, например, методом случайных испытаний (методом бутстрапа).

2. Нерешенные ранее части поставленной проблемы

Ещё раз отметим, что существенному расширению возможностей статистических методов РИК способствовало применение СМВР для вероятностного описания признаков-характеристик изучаемых объектов. Рассмотрению СМВР к настоящему времени посвящено большое количество публикаций [4-12 и другие]. Во всех этих работах в основном рассматриваются полигауссовские вероятностные модели с достоверной (а не с вероятностной) принадлежностью классам объектов в ОВ.

При этом выбор распознающих признаков объектов без классификации ограничивается преобразованием Карунена-Лоэва исходных признаков и последующей отбраковкой зашумлённых вторичных признаков с малой дисперсией. Недостаточность такого сокращения размерности пространства распознающих признаков общеизвестна [3,5,7].

В книге одного из авторов этой статьи [7] и, насколько нам известно, только в ней указывается на необходимость расчёта НДГ в БВ. НДГ в БВ необходима для обоснования надёжности классификации каждого предъявляемого объекта и для построения несмещенного ККК с целью его оптимизации без привлечения дополнительной контрольной выборки помимо ОВ. Однако в этой работе [7] *информационная матрица Фишера* (ИМФ) вычисляется только в предположении достоверной принадлежности определённым классам объектов ОВ.

3. Цель работы

Цель работы – создание нового универсального субоптимального метода РИК для содержательной обработки и автоматического дешифрирования дистанционных данных, точнее, данных *дистанционного зондирования* (ДЗ) на СМВР признаков-характеристик изучаемых объектов.

Данные ДЗ подлежат содержательной обработке статистическими методами по трём причинам. Во-первых, наблюдения ДЗ имеют повышенную флуктуативность в силу естественных и технических условий их получения. Во-вторых, развитые в настоящее время достаточно гибкие на СМВР статистические методы способны решать новые сложные задачи РИК в рамках класса вероятностных моделей, адекватных практическим наблюдениям ДЗ. В-третьих, статистические модели и методы опираются на математические обоснования, обеспечивающие этим моделям и методам надёжный ККК.

Предлагаемый метод РИК обладает следующими новационными свойствами.

Проведена оптимизация количества распознающих признаков-характеристик классифицируемых объектов на предварительной стадии обработки данных (без выполнения классификации) и на основе ККК (с выполнением классификации) [7,10].

ОВ объектов, используемая в качестве входа в метод РИК, допускается вероятностной, а не достоверной относительно принадлежности объектов этой выборки заданным классам [7,10]. Такое свойство ОВ имеет три важных для практики следствия: во-первых, информационные объёмы ОВ могут во много раз возрасти; во-вторых, единым алгоритмом решаются три различных задачи РИК – распознавание с обучением, распознавание с частичным обучением и распознавание с самообучением; в-третьих, объекты ОВ могут быть элементами нечётких множеств.

Заданное семейство ЗРВ признаков для объектов каждого класса является функционально достаточно полным для аппроксимации с необходимой точностью истинного распределения признаков при практическом решении задач РИК. Таким свойством обладает, например, полигауссова СМВР [7,9].

Построенное на основе ККК распределение признаков-характеристик из заданного семейства ЗРВ имеет минимальное (либо близкое к минимальному) количество оцениваемых при обучении параметров, что позволяет сократить необходимый объём репрезентативной ОВ [7,9,10].

В частности, при использовании в качестве семейства ЗРВ смесевых моделей распределений выбирается СМВР с минимально возможным числом компонент, решающая задачу РИК с заданной точностью [7,9,10].

При проведении оптической, тепловой, радиоволновой, поли-, гиперспектральной и других съёмок для широкого разнесения во времени этапов обучения и распознавания применяется *динамическая модель распознавания* (ДМР). Как показывает практика проведение некоторых дистанционных работ, обучение и распознавание с применением ДМР разносятся во времени не на минуты и часы, а на месяцы и даже годы [7,13,14,15].

Итог работы метода РИК осуществляется нахождением не (как принято) БВ принадлежности предъявляемых объектов заданным классам, а расчётом НДГ для БВ. Оценка НДГ в БВ производится при помощи ИМФ [7,16,17].

Субоптимальность метода РИК реализуется максимизацией ККК, который представляет собой взвешенную сумму НДГ для БВ принадлежности собственным классам объектов ОВ. Максимизация ККК производится на подмножествах признаков-характеристик объектов ОВ и на количествах компонент в СМВР характеристик по всем заданным классам. Количеством компонент во всех СМВР классов определяется оптимальное правило классификации [7].

4. Основные составляющие блоки нового метода РИК для обработки данных ДЗ

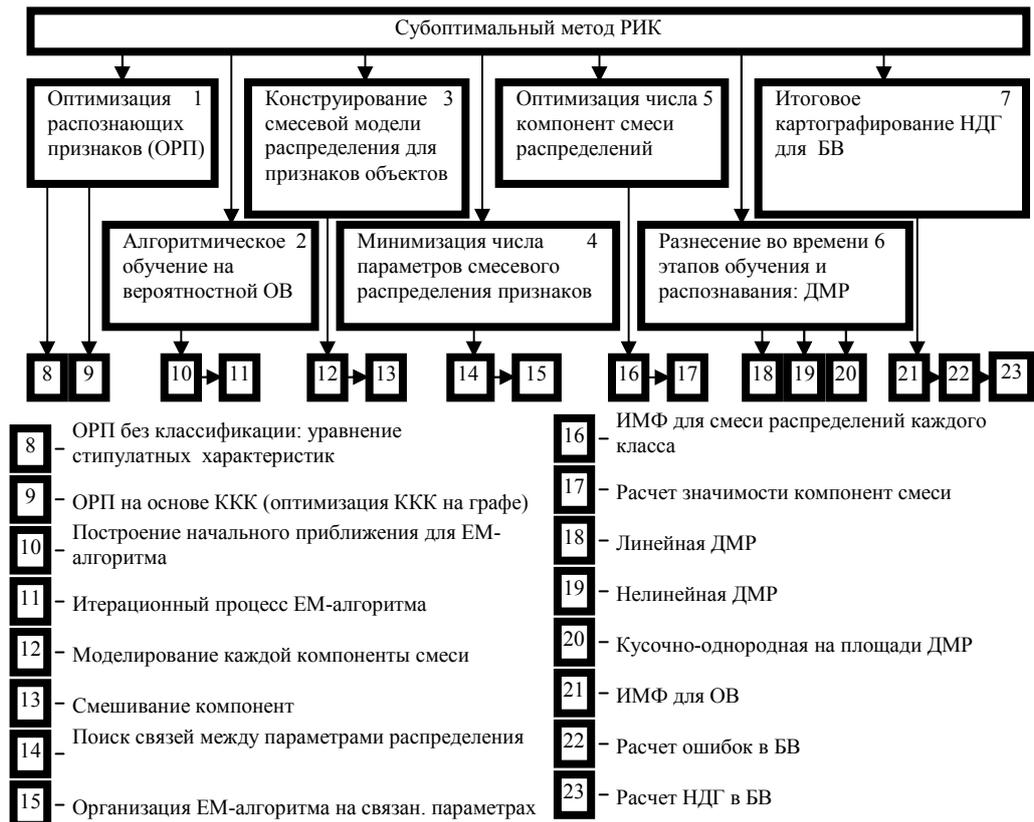


Схема 1. Субоптимальный метод РИК

На схеме показаны основные блоки субоптимального метода РИК.

Как выше указано в данной работе (3. Цель работы), часть блоков этой схемы принципиально выполнена (см. ссылки [7,10,12-15]). Рассмотрим далее три блока: **8** – ОРП без классификации; **11** – итерационный процесс EM-алгоритма; **21** – ИМФ на вероятностной ОВ и её применение в субоптимальном методе РИК. Именно эти блоки включают новый либо существенно обновленный математический аппарат.

4.1. ОРП без классификации

Табл.1. Термины и обозначения для ОРП

Термины	Обозначения
Матрица числовых признаков однородных объектов (исходные данные)	$Z = \{z_{ik}\}$
Множество признаков индексов	$I = (i = 1, 2, \dots, m)$
Множество объектных индексов	$K = (k = 1, 2, \dots, n); n \gg m$
Объект: перечень упорядоченных поименованных числовых признаков	$\bigvee_{k=1}^n [z_{1k}, z_{2k}, \dots, z_{mk}]$
Множество однородных объектов: множество объектов с одинаково упорядоченными одноименными признаками	$(z_{11}, z_{21}, \dots, z_{m1}); (z_{12}, z_{22}, \dots, z_{m2})$ $(z_{1n}, z_{2n}, \dots, z_{mn})$
Граундкалибровочное преобразование признаков однородных объектов	$Y = G_I [Z] = G(g_{1i}, \dots, g_{qi}; Z)$ $Y = \{y_{ik}\}$
Граундкалибровочные параметры (коэффициенты) признаков	$\bigvee_{i=1}^m [g_{1i}, g_{2i}, \dots, g_{qi}]$
Граундфакторы признаков	$\bigvee_{i=1}^m [X_i = (x_{1i}, x_{2i}, \dots, x_{ri})]$
Компоненты граундфакторов могут быть функционально взаимосвязаны общими параметрами U_i	$X_i = F_i(U_i); U_i = \{u_{ij}\}; (j = 1, \dots, s);$ $x_{li} = f_{li}(u_{ij}); (l = 1, 2, \dots, r); s \leq r$
Стипулатные характеристики объектов	$\bigvee_{k=1}^n [A_k = (a_{1k}, a_{2k}, \dots, a_{pk})]; p \ll m$
Объектнохарактеристическое преобразование граундфакторов	$Y = H_k [X_i] = H(A_k; X_i); Y = \{y_{ik}\}$
Фундаментальное уравнение стипулатных характеристик	$G_I [Z] = H_k [X_I]$

Математическая постановка задачи

Пусть задан вид граундкалибровочного преобразования признаков $G_I [Z]$ и задан вид объектно-характеристического преобразования граундфакторов $H_k [X_i]$. Найти стипулатные характеристики объектов A_k , а также граундкалибровочные параметры g_{1i}, \dots, g_{qi} и граундфакторы X_i признаков, удовлетворяющие фундаментальному уравнению:

$$\exists_G \exists_H \{ \bigvee_{k=1}^n \bigvee_{i=1}^m [G(g_{1i}, \dots, g_{qi}; Z) = H(A_k; X_i)] \} (*)$$

Система (*) содержит (nxt) уравнений относительно $[np + m(q + r)]$ неизвестных. Пример уравнения стипулатных характеристик (*), когда преобразования G_I и H_K являются линейными операторами относительно аргументов Z и X_i ; при этом $q = 2; r = p - 1$;

$$\forall_{k=1}^n \forall_{i=1}^m [g_{0i} + g_{1i} z_{ik} = a_{0k} + a_{1k} x_{1i} + a_{2k} x_{2i} + \dots + a_{p-1k} x_{p-1i}] (**)$$

Это система (nxt) нелинейных уравнений относительно $((p + 1)m + pn)$ неизвестных $g_{0i}, g_{1i}, x_{1i}, x_{2i}, \dots, x_{p-1i}, a_{0k}, a_{1k}, a_{2k}, \dots, a_{p-1k}$.

Для получения единственного решения системы уравнений (**) на граундкалибровочные коэффициенты g_{0i}, g_{1i} и граундфакторы X_i признаков накладываются дополнительные (опорные) ограничения. С этой же целью ограничительными соотношениями должны быть связаны размерности n, m, p, q, r, s, v всех переменных в уравнении (*).

4.2. Итерационный процесс EM-алгоритма на вероятностной ОВ

Решение уравнения правдоподобия для вероятностной обучающей выборки сводится к новой модифицированной версии EM-алгоритма.

В схеме EM-алгоритма использованы следующие обозначения: $\{z^i\}_{i=1}^n$ - множество объектов обучающей выборки; $\{z_k^i\}_{k=1}^m$ - множество признаков i -го объекта; $\{\rho_s^i\}_{s=1}^Q$ - вектор вероятностей принадлежности i -го объекта различным классам; $\forall i \sum_{s=1}^Q \rho_s^i = 1$; Q - число классов; $\{q_r^s\}_{r=1}^{V_s}$ - вектор вероятностей принадлежности объектов обучающей выборки различным компонентам s -го класса; $\forall s \sum_{r=1}^{V_s} q_r^s = 1$; V_s - число компонент в классе с номером s ; $N(z, G, A) = \frac{1}{\sqrt{(2\pi)^m |G|}} \exp\left[-\frac{1}{2}(z - A)^T G^{-1}(z - A)\right]$ - гауссова плотность распределения вектора z с матрицей ковариации G и вектором математических ожиданий A . $G^{-1} = \left\{ \begin{matrix} (-1) \\ g_{k_1 k_2} \end{matrix} \right\}$ - матрица, обратная к матрице ковариаций.

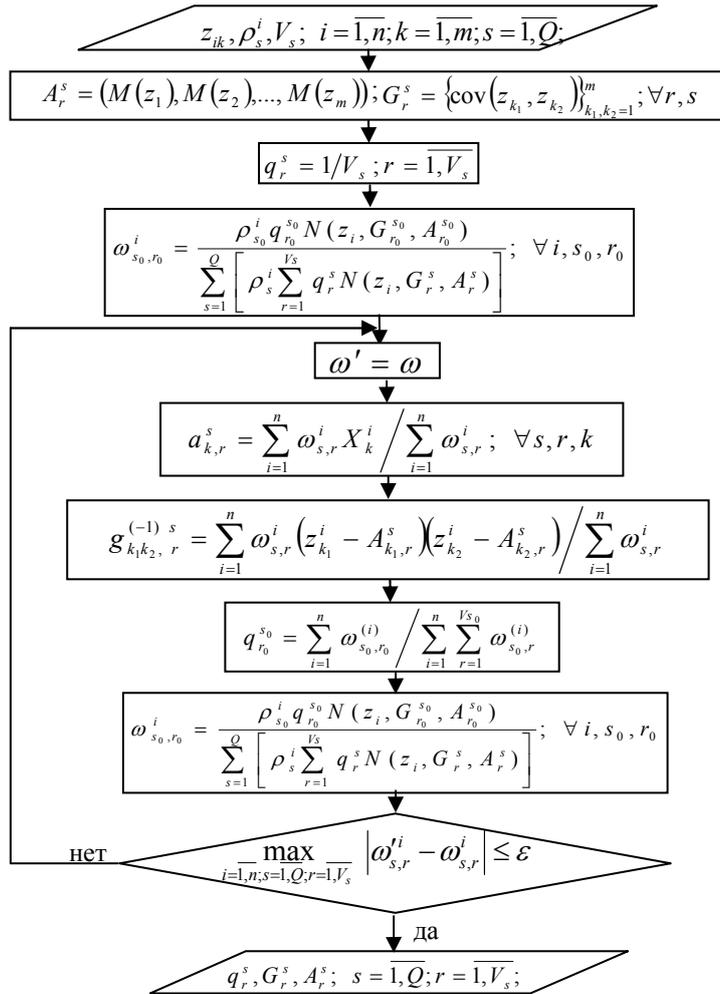


Схема 2. Итерационный процесс EM-алгоритма на вероятностной ОБ

4.3. ИМФ для вероятностной ОБ и ее применение в субоптимальном методе РИК

Пусть последовательность наблюдаемых случайных величин $x = (x_1, x_2, \dots, x_n)$ имеет плотность распределения вероятностей $f(x, \alpha)$, где $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m)$ - вектор неизвестных параметров распределения. Тогда ИМФ для этих наблюдений X относительно параметров α имеет вид[16]:

$$I = \int_{B(X)} \text{grad}(\ln f) \alpha \text{grad}^T(\ln f) \alpha f(X, \alpha) dX$$

где $B(X) = \{X : f(X, \alpha) \neq 0\}$

Если наблюдения $x = (x_1, x_2, \dots, x_n)$ являются статистически независимыми, то

$$f(X, \alpha) = \prod_{i=1}^n f_i(x_i, \alpha)$$

где $f_i(x_i, \alpha)$ - плотности распределения наблюдений x_i .

Тогда имеет место равенство:

$$I = \sum_{i=1}^n \int_{b(x_i)} \underset{\alpha}{grad}(\ln f_i) \underset{\alpha}{grad}^T(\ln f_i) f_i(x_i, \alpha) dx_i,$$

где $b(x_i) = \{x_i : f_i \neq 0\}$

В частном случае, когда $\forall_{i=1}^n [f_i(x_i, \alpha) = f(x_i, \alpha)]$, то есть наблюдения X_i равновероятностные, справедливо:

$$I = n \int_{b(x_i)} \underset{\alpha}{grad}(\ln f) \underset{\alpha}{grad}^T(\ln f) f(x_i, \alpha) dx_i$$

Пусть вероятностная ОВ имеет S различных зон распознавания, в каждой из которых находится (n_1, n_2, \dots, n_s) объектов. Пусть зона с номером s имеет плотность распределения объектов f_s . При статистической независимости всех наблюдаемых объектов x_i получаем равенство:

$$I = I_1 + I_2 + \dots + I_S$$

где $I_s = n_s \int_{b(x_{i_s})} \underset{\alpha}{grad}(\ln f_s) \underset{\alpha}{grad}^T(\ln f_s) f_s(x_{i_s}, \alpha) dx_{i_s}$

$$b(x_{i_s}) = \{x_{i_s}, \alpha \neq 0\}$$

Если $\hat{\alpha}$ - оценка, полученная *методом максимального правдоподобия* (ММП), то неравенство Крамера-Рао в матричном виде будет таким: $K(\hat{\alpha}) \geq I^{-1}$, а при $n \rightarrow \infty$ справедлива асимптотика $I^{-1} \approx K(\hat{\alpha})$ [17], где $K(\hat{\alpha})$ - ковариационная матрица оценок ММП. На её диагонали стоят квадраты ошибок всех параметров СМВР.

Для расчёта ошибок ΔP в байесовских вероятностях $P(\alpha)$ используется формула переноса ошибок:

$$\Delta^2 P(\alpha) = \underset{\alpha}{grad}^T P * K(\alpha) * \underset{\alpha}{grad} P$$

Чтобы получить НДГ в БВ, необходимо из БВ вычесть соответствующие им ошибки с нормировочным коэффициентом надёжности.

4.4. Оптимизация статистического метода РИК на СМВР.

БВ принадлежности объекта z^i классу с номером s_0

$$(s_0 = 1, 2, \dots, Q) : P(s_0 | z^i) = \frac{\pi_{s_0} f_{s_0}(z^i)}{\sum_{s=1}^Q \pi_s f_s(z^i)}; \sum_{s=1}^Q \pi_s = 1 \quad (1)$$

$$\text{НДГ в БВ обозначим } R(s_0, z^i) : R(s_0, z^i) = P(s_0 | z^i) - \gamma \cdot \Delta P(s_0 | z^i) \quad (2)$$

Здесь γ – коэффициент надежности относительно принадлежности объекта z^i классу с номером s_0 ; $P(s_0 | z^i)$ – ошибка в расчете БВ, вычисляемая с помощью ИМФ и зависящая от объема n ОБ: (z^1, z^2, \dots, z^n) .

Введем ККК:

$$K = \frac{1}{n} \sum_{i=1}^n R(s_{0i}, z^i), \quad (3)$$

где s_{0i} – номер класса, на котором $R(s_0, z^i)$ для объекта z^i с номером i (1), (2) принимают максимальные значения.

Правило классификации (нерандомизированное) для объекта z^i :

$$R(s_0, z^i) \geq R(s, z^i) \quad \forall s = \overline{1, Q} \Rightarrow z^i \in s_0$$

Это правило определяется числом компонент $\{V_s\}_{s=1}^Q$ в СМВР для каждого класса, где параметры плотности являются оценками ММП.

Статистический метод классификации, на котором ККК (3) становится максимальным по числу компонент $\{V_s\}_{s=1}^Q$ и по набору распознающих признаков m_0 , назовем *субоптимальным статистическим методом* (СОСМ).

Для СОСМ выполняется:

$$K_{opt} = \max_{m_0, \{V_s\}_{s=1}^Q} K(m_0, \{V_s\}_{s=1}^Q).$$

5. Работа программного комплекса

А. Создание имитационного примера:

- 1) Моделирование объектно-информационной карты
- 2) Моделирование ковариационных эллипсов (эллипсоидов)
- 3) Моделирование стипулатных характеристик всех объектов
- 4) Моделирование параметров признаков: граундкалибровочных параметров и граундфакторов

5) Моделирование признаков по стипулатным характеристикам объектов

В. Задача классификации:

1. Уменьшение размерности признаков пространства:

1.1. Выделение необходимого числа областей для получения средних значений признаков в этих областях

1.2. Моделирование необходимого числа стипулатных характеристик для нахождения всех параметров признаков

1.3. Нахождение характеристик объектов на выделенных областях

1.4. Итерационный процесс нахождения оптимальных характеристик объектов

2. EM-алгоритм:

2.1. Вычисление характеристик объектов ОВ

2.2. Нахождение начальных ковариационных эллипсов (эллипсоидов)

2.3. EM-алгоритм нахождения неизвестных параметров СМВР

2.4. Нахождение характеристик всех объектов на объектно-информационной карте

2.5. Расчёт байесовских вероятностей принадлежности объектов заданным классам

Работу программного комплекса можно разделить на две части: решение прямой задачи (А) и решение обратной задачи (В).

А. Входным параметром для разработанного программного комплекса является объектно-информационная карта, на которой цвет объекта отвечает за его вероятностную принадлежность тому или иному классу. Далее следует моделирование смесевой модели распределения классов. Вектора средних определяют положения центров эллипсов (эллипсоидов), а матрицы ковариаций – длины и повороты осей эллипсов (эллипсоидов). Таким образом, с помощью сгенерированных векторов средних и матриц ковариаций можно определить положение эллипса (эллипсоида) в плоскости (пространстве). В зависимости от необходимого количества характеристик происходит создание характеристик всех объектов. Создание граундфакторов признаков напрямую зависит от количества смоделированных характеристик. У каждого объекта (***) количество граундфакторов меньше на единицу, чем количество его характеристик. Следующим этапом было моделирование калибровочных параметров признаков. Имея характеристики объектов, граундфакторы и калибровочные параметры признаков, можно сформировать числовые признаки. На этом решение прямой задачи, а именно создание имитационного примера, заканчивается.

В. Решение обратной задачи начинается с нахождения оптимального количества стипулатных характеристик объектов. На объектно-информационной карте выделяется несколько областей (их количество зависит от модели, которая ищется). Если на объектно-информационной карте, как в имитационных примерах, имеется обучающая выборка, то удобно брать области из разных классов. На этих областях находится среднее значение признаков. Для решения системы уравнений, неизвестными в которых будут параметры признаков, необходимо смоделировать некоторое число характеристик (которое опять же зависит от искомой модели). Эти характеристики не могут быть произвольными:

на них накладываются ограничения, чтобы решаемая система была совместной. По найденным параметрам признаков осуществляется поиск всех характеристик объектов. Найденные характеристики являются начальным приближением для итерационного процесса, который следует далее. Характеристики, найденные итерационным процессом, используются для нахождения начальных приближений EM-алгоритма, а именно: ковариационные эллипсы (эллипсоиды) и координаты их центров.

Полученные начальные приближения подаются в модуль программы, реализующий EM-алгоритм, который находит неизвестные параметры смеси для объектов обучающей выборки. После нахождения характеристик всех объектов, программа приступает к распознаванию предъявляемых объектов путем подсчета байесовских вероятностей принадлежности объектов заданным классам.

6. Результаты работы программного комплекса

Далее приведена иллюстрация работы имитационного примера. В данном примере участвуют три класса (условно обозначим их: «первый», «второй» и «третий»). На рис.1.1,1.2,1.3 представлены карты вероятностной принадлежности объектов тому или иному классу. В зависимости от цвета объекта можно определить его вероятность принадлежности классу (рис.1.4) Общее число объектов – 86800, из них в обучающей выборке – 1380.

В «первом» и «третьем» классах по две компоненты распределения, во «втором» - 3. Собственные векторы и собственные числа матрицы ковариации каждой из компонент позволяют оценить размеры и форму облака распределения случайной величины, аппроксимировав его ковариационным эллипсоидом (эллипсом в двумерном случае) (рис.2).

По данным распределениям были созданы характеристики для каждого из объектов.

Далее, для каждого признака были сгенерированы граундфакторы и калибровочные параметры. С их помощью были созданы 70 признаков для каждого из объектов (рис. 3). Решение прямой задачи закончено.

Решение обратной задачи начинается с уменьшения признакового пространства. Для этой цели случайным образом создается выборка из общего множества объектов. В нашем случае 5% от общего числа объектов, т.е. 4340. Далее находятся характеристики объектов и определяются параметры признаков. Для их определения было решено в данном примере 303800 уравнений с общим числом в 13300 неизвестных.

В результате получаем пространство характеристик нужной размерности (в данном случае - 3).

По полученным граундфакторам и граундкалибровочным параметрам определяются характеристики объектов обучающей выборки. Результатом работы EM-алгоритма являются, как уже было сказано, весовые вероятностные коэффициенты для каждой компоненты, вектора средних значений компонент и ковариационные матрицы компонент каждого класса. По этим данным определяются байесовские вероятности принадлежности объектов каждому из классов (рис.4.1,4.2,4.3).

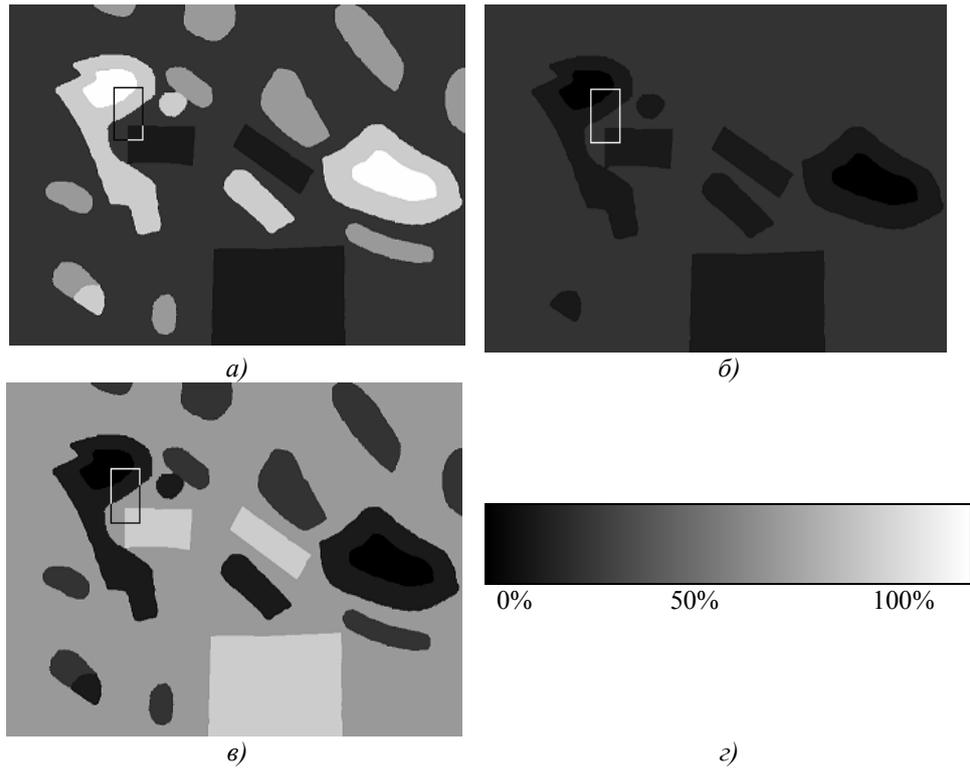


Рис.1 Вероятностные объектно-информационные карты трех классов с выделенной обучающей выборкой: а) - «первый» класс, б) - «второй» класс, в) - «третий» класс, г) - цвета объекта

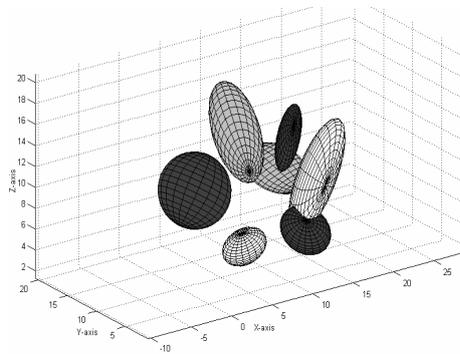


Рис.2. Ковариационные эллипсоиды всех семи компонент классов.

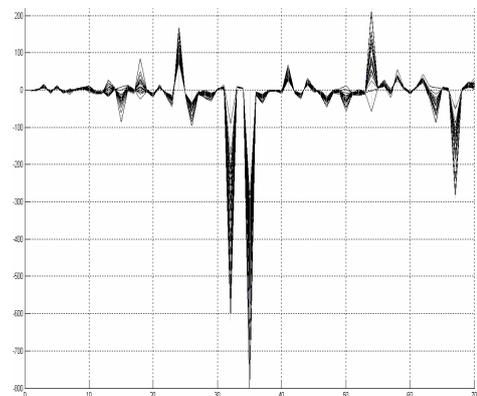


Рис.3. Семейство спектрограмм 20-ти объектов.

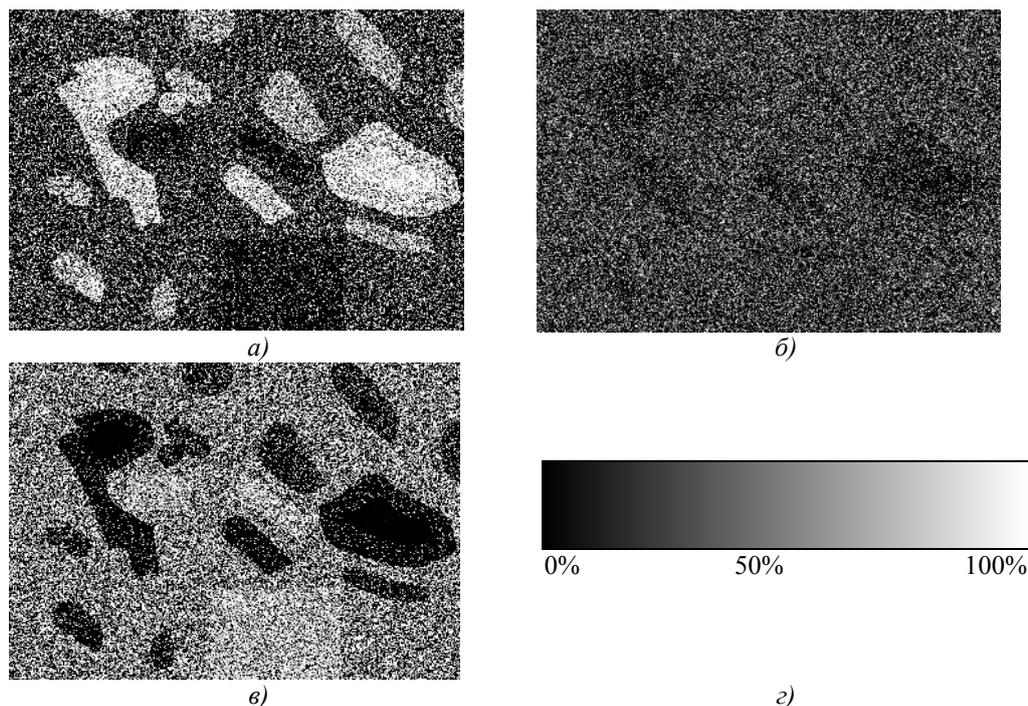


Рис. 4. Байєсовские вероятности принадлежности объектов каждому из классов: а) - «первый» класс, б) - «второй» класс, в) - «третий» класс, г) - цвета объекта

7. Выводы и перспективы дальнейших исследований

На принципиальном уровне построения единой математической модели субоптимальный метод РИК для содержательной обработки данных ДЗ на СМВР полностью реализован. Однако его усовершенствование и развитие и в математическом моделировании, и в компьютерном программировании необходимо. Это относится ко всем блокам схемы субоптимального метода РИК, но прежде всего, нужно:

- оптимизировать без классификации число стипулатных характеристик объектов;
- усовершенствовать метод нахождения стипулатных характеристик объектов, включая возможность нелинейного преобразования исходных признаков и учёт нелинейных связей граундфакторов признаков;
- выйти за рамки полигауссовых СМВР для признаков-характеристик объектов;
- минимизировать число параметров СМВР с учётом возможных зависимостей между этими параметрами;
- усовершенствовать построение нелинейной и кусочно-однородной ДМР для разнесения во времени этапов обучения и распознавания;
- с целью сокращения времени работы компьютера усовершенствовать вычислительные алгоритмы всех итерационных процессов: ОРП, ЕМ-алгоритма, построения ИМФ, ДМР, ККК, оптимизации метода РИК;
- на основе программного комплекса РИК в системе Matlab создать программный продукт РИК для содержательной обработки и автоматического

дешифрирования данных ДЗ, пригодный широкому кругу пользователей независимо от их специальности.

ЛИТЕРАТУРА

1. Журавлев Ю.И., Рязанов В.В., Сенько О.В. — Распознавание. Математические методы. Программная система. Практические применения. М.: Фазис, 2006.
2. Соловьев Н. В. Обработка и распознавание изображений в системах превентивной безопасности. Санкт-Петербург, 2005. – 153 с.
3. Верхаген, К., Дейн, Р. И др. Распознавание образов: состояние и перспективы. – М.: Радио и связь, 1985 – 104 с.
4. Дуда, Р., Харт, П. Распознавание образов и анализ сцен. – М.: Мир, 1976. – 559 с.
5. Патрик Э.А. Основы теории распознавания образов.–М.: Сов.радио, 1980. – 208 с.
6. С.А. Айвазян, З.Н. Бежаева, О.В. Староверов Классификация многомерных наблюдений. – М.: Статистика, 1974. – 238 с.
7. Колтунов И.А. Применение смесевых моделей вероятностных распределений для обработки изображений и распознавания образов. – Белгород: БелГУ, 2004. – 122 с.
8. G.J. McLachlan and K.E. Basford, Mixture Models: Inference and Applications to clustering, New York: Marcel Dekker, 1988.
9. G.J. McLachlan and T. Krishnan, The EM-Algorithm and Extensions. New York: Wiley, 1997.
10. Koltunov, A. and Ben-Dor, E. Mixture density separation as a tool for high-quality interpretation of multi-source remote sensing data and related issues. International Journal of Remote Sensing, v. 25, pp.3275-3299, 2004.
11. Трофимов А.Т. Полигауссовские вероятностные модели и синтез информационных систем. – В. Новгород: НГУ. 2002, – 183 с.
12. Y. Koltunov, et al. Detection and recognition of objects by multispectral sensing. US Patent, 6837617, January, 2005.
13. Koltunov, J., Koltunov, A. Dynamic Detection Model and its application for perimeter security, intruder detection, and automated target. In: Infrared Technology and Applications XXIX, Proc. SPIE, v. 5074, pp. 777-787, 2003.
14. A. Koltunov, S.L. Ustin. Early fire detection using non-linear multitemporal prediction of thermal imagery. Remote Sensing of Environment, 110(2007),18-28.
15. A. Koltunov, E. Ben-Dor, S.L. Ustin. Image construction using multitemporal observations and Dynamic Detection Models. International Journal of Remote Sensing, vol. 30, №1, January 2009, pp. 57-83.
16. Рао С.Р. Линейные статистические методы и их применения. – М.: Наука, 1968 – 548 с.
17. Колтунов И.А. О статистиках с минимальными дисперсиями «Теория вероятностей и её применения». - 1977. - №3. - С.642-644.