

РАНДОМИЗАЦИЯ МНОГОМЕРНЫХ РАСПРЕДЕЛЕНИЙ  
В МЕТРИКЕ МАХАЛАНОВИСА

© 2000 г. Л. В. Лабунец

Поступила в редакцию 30.05.2000 г.

Представлен модифицированный *EM*-алгоритм оценивания параметров эллипсоидально-симметричных распределений. Алгоритм адаптирован как к выборочной, так и к регулярной формам представления исходной информации. Реализация *w*-взвешенных (в смысле Л.Д. Мешалкина) оценок параметров парциальных распределений позволяет алгоритму эффективно работать с малыми выборками, малыми межклассовыми расстояниями и вырожденными распределениями. Методами цифрового моделирования исследованы свойства сходимости алгоритма для различных функционалов качества обучения. Показано, что критерий максимального правдоподобия обеспечивает большую скорость сходимости по сравнению с функционалами расстояния Бхатачария и дивергенции Кульбака.

## ВВЕДЕНИЕ

Параметрическое описание многомерных распределений в виде конечных смесей стандартных (парциальных) плотностей является фундаментальной моделью для ряда разделов прикладной статистики, таких как статистическое моделирование, выбросы случайных процессов, прогнозирование, распознавание и классификация. В обширной литературе по этим предметам отметим здесь монографии [1–5].

Широко известный *EM*-алгоритм расщепления смесей вероятностных распределений [6] ориентирован на выборочную форму представления информации. В этом случае исходная *N*-мерная плотность распределения вероятностей  $\phi(x_1, \dots, x_N)$  не определена явно, но задана случайная независимая выборка  $(X_1^{(m)}, \dots, X_N^{(m)})$  ( $m = 1, \dots, M$ ), распределенная согласно  $\phi(x_1, \dots, x_N)$ . Однако в практических приложениях информация о многомерной плотности, не обязательно вероятностей, может быть представлена в регулярной форме, под которой будем понимать множество значений распределения, заданных в узлах регулярного растра. В такой ситуации исследователь оказывается при решении, например, задач статистического моделирования случайного нестандартного вектора, оценки многомерного интеграла или спектрального оценивания сигналов. Очевидно, что параметрическая модель в виде смеси парциальных распределений и алгоритм идентификации ее параметров, адаптированный к регулярной форме исходной информации, весьма плодотворны для решения подобного рода задач.

В последнее время широкое распространение получили полупараметрические оценки вероятностных распределений [7, 8]. Представляя собой

разумное сочетание параметрических и непараметрических моделей, такого рода оценки являются мощным инструментарием описания многомерных вероятностных распределений. В этой связи естественным представляется стремление адаптировать алгоритм идентификации параметров конечной смеси распределений к ее полупараметрической форме.

Данная работа посвящена рассмотрению возможных вариантов *EM*-алгоритма расщепления смесей эллипсоидально симметричных распределений с точки зрения альтернативных форм представления исходной информации – выборочной или регулярной, а также с точки зрения полупараметрической модели многомерного распределения.

## 1. ПОСТАНОВКА ЗАДАЧИ

В качестве модели структуры *N*-мерных данных рассмотрим кластерную, т.е. совокупность *K* “сгустков” случайных точек (в случае выборочной формы) или *K* модальных значений исходного распределения  $\phi(\vec{X})$  (в случае регулярной). Число *K* парциальных плотностей распределений  $\hat{\phi}_k(\vec{X}|\vec{\theta})$  в смеси

$$\hat{\phi}(\vec{X}|\vec{\theta}) = \sum_{k=1}^K p_k \hat{\phi}_k(\vec{X}|\vec{\theta}_k); \quad (1)$$

$$(\vec{X} = (x_1, \dots, x_N)^T) \in R^N,$$

а также начальное приближение вектора параметров смеси  $\vec{\theta}^{(0)}$  считаем известными и полученными на этапе разведочного анализа данных, на-

пример, методами целенаправленного проецирования [9] или динамической визуализации [10]. Кластерную структуру удобно описывать эллипсоидально-симметричными плотностями

$$\hat{\phi}_k(\vec{X}|\vec{\theta}_k) = S_1^{-1} v_{N-1}^{-1} (\det \mathbf{B}_k)^{-\frac{1}{2}} f(D_0^2(\vec{X}|\vec{\theta}_k)). \quad (2)$$

Здесь  $S_1 = (2\pi^{N/2})\Gamma(N/2)$  – площадь поверхности единичной сферы в  $R^N$ ;  $D_0^2(\vec{X}|\vec{\theta}_k) = (\vec{X} - \vec{M}_k)^T \mathbf{B}_k^{-1} (\vec{X} - \vec{M}_k)$  – расстояние Махаланобиса; индекс “т” – символ транспонирования;  $f(y^2)$  – одномерное, монотонно убывающее при  $y \rightarrow \infty$  распределение с конечным  $(N - 1)$ -м моментом  $v_{N-1} = \int_0^\infty \rho^{N-1} f(\rho^2) d\rho$ ;  $\vec{\theta} = (p_1, \dots, p_K, \vec{\theta}_1, \dots, \vec{\theta}_K)$  – вектор параметров смеси, компоненты которого  $p_k$  – веса (априорные вероятности), а  $\vec{\theta}_k = (\vec{M}_k, \mathbf{B}_k)$  – характеристики положения и масштаба парциальных распределений ( $k = 1, \dots, K$ ).

Целью обучения в задаче идентификации смеси (1) будем считать поиск значений вектора параметров  $\vec{\theta}^*$ , обеспечивающих экстремумы меры различия между исходным  $\phi(\vec{X})$  и модельным  $\hat{\phi}(\vec{X}|\vec{\theta})$  распределениями. Оптимальными статистическими свойствами обладает функционал правдоподобия Фишера

$$L(\vec{\theta}) = \int \phi(\vec{X}) \ln \hat{\phi}(\vec{X}|\vec{\theta}) d\vec{X}, \quad (3)$$

$$(d\vec{X} = dx_1 dx_2 \dots dx_N).$$

При достаточно широких предположениях относительно характера поведения меры различия (3) максимально правдоподобные оценки параметров смеси (1) вида  $\vec{\theta}^* = \arg \max \{L(\vec{\theta})\}$  состоятельны и асимптотически эффективны в рамках выборочного варианта EM-алгоритма [11]. Однако правдоподобие не единственная мера различия с хорошими статистическими свойствами в контексте проблемы оценки параметров смеси распределений. На наш взгляд заслуживают внимания функционалы расстояния Бхатачария и дивергенции Кульбака

$$D(\vec{\theta}) = -\ln \int [\phi(\vec{X}) \hat{\phi}(\vec{X}|\vec{\theta})]^{-\frac{1}{2}} d\vec{X}; \quad (4)$$

$$J(\vec{\theta}) = \int [\phi(\vec{X}) - \hat{\phi}(\vec{X}|\vec{\theta})] \ln \frac{\phi(\vec{X})}{\hat{\phi}(\vec{X}|\vec{\theta})} d\vec{X}$$

и соответствующие им оптимальные оценки параметров смеси вида  $\vec{\theta}^* = \arg \min \{D(\vec{\theta})\}$ ;  $\vec{\theta}^* = \arg \min \{J(\vec{\theta})\}$ .

## 2. СИСТЕМА НОРМАЛЬНЫХ УРАВНЕНИЙ

Дифференцируя функционалы  $L(\vec{\theta}) + \lambda(p_1 + \dots + p_K)$ ;  $D(\vec{\theta}) + \lambda(p_1 + \dots + p_K)$  и  $J(\vec{\theta}) + \lambda(p_1 + \dots + p_K)$  по составляющим вектора  $\vec{\theta}$ , где  $\lambda$  – множитель Лагранжа, и приравнявая результат нулю, после громоздких преобразований, аналогичных [12], с учетом ограничения  $p_1 + \dots + p_K = 1$ , получим систему нормальных уравнений относительно параметров смеси парциальных распределений:

$$\begin{cases} \int [p_k - \text{Pr}(k|\vec{X}, \vec{\theta})] w(\vec{X}|\vec{\theta}) d\vec{X} = 0, \\ \int (\vec{X} - \vec{M}_k) F\{D_0^2(\vec{X}|\vec{\theta}_k)\} \text{Pr}(k|\vec{X}, \vec{\theta}) w(\vec{X}|\vec{\theta}) d\vec{X} = 0, \\ \int [F\{D_0^2(\vec{X}|\vec{\theta}_k)\} (\vec{X} - \vec{M}_k) (\vec{X} - \vec{M}_k)^T + \\ + \vec{B}_k/2] \text{Pr}(k|\vec{X}, \vec{\theta}) w(\vec{X}|\vec{\theta}) d\vec{X} = 0 \end{cases}$$

( $k = 1, 2, \dots, K$ ).

Здесь  $\text{Pr}(k|\vec{X}, \vec{\theta}) = p_k \hat{\phi}_k(\vec{X}|\vec{\theta}_k) / \hat{\phi}(\vec{X}|\vec{\theta})$  – так называемые *апостериорные распределения* принадлежности наблюдения  $\vec{X}$  к  $k$ -му классу. Очевидно, что  $\text{Pr}(k|\vec{X}, \vec{\theta}) \geq 0$  и  $\text{Pr}(1|\vec{X}, \vec{\theta}) + \dots + \text{Pr}(K|\vec{X}, \vec{\theta}) = 1$  для любого  $\vec{X}$  из области определения смеси;  $F(\rho^2) = f'_\rho(\rho^2)/f(\rho^2)$ ;  $w(\vec{X}|\vec{\theta})$  – *весовая функция*, определяемая типом функционала качества оценок параметров смеси: для правдоподобия –  $w(\vec{X}|\vec{\theta}) = \phi(\vec{X})$ ; для расстояния –  $w(\vec{X}|\vec{\theta}) = \{\phi(\vec{X}) \hat{\phi}(\vec{X}|\vec{\theta})\}^{1/2}$ ; для дивергенции –  $w(\vec{X}|\vec{\theta}) = \phi(\vec{X}) - \hat{\phi}(\vec{X}|\vec{\theta}) \{1 - \ln[\phi(\vec{X})/\hat{\phi}(\vec{X}|\vec{\theta})]\}$ .

Нетрудно видеть, что приведенная система нелинейных уравнений допускает преобразование к каноническому виду. Естественно пытаться получить ее решение с помощью простейшего численного метода последовательных приближений:

$$p_k^{(i+1)} = \frac{\int \text{Pr}(k|\vec{X}, \vec{\theta}^{(i)}) w(\vec{X}|\vec{\theta}^{(i)}) d\vec{X}}{\int w(\vec{X}|\vec{\theta}^{(i)}) d\vec{X}}; \quad (5)$$

( $k = 1, 2, \dots, K$ )

$$\vec{M}_k^{(i+1)} = \frac{\int \vec{X} F\{D_0^2(\vec{X}|\vec{\theta}_k^{(i)})\} \Pr(k|\vec{X}, \vec{\theta}^{(i)}) w(\vec{X}|\vec{\theta}^{(i)}) d\vec{X}}{\int F\{D_0^2(\vec{X}|\vec{\theta}_k^{(i)})\} \Pr(k|\vec{X}, \vec{\theta}^{(i)}) w(\vec{X}|\vec{\theta}^{(i)}) d\vec{X}}; \quad (6)$$

$$\mathbf{B}_k^{(i+1)} = -2 \frac{\int (\vec{X} - \vec{M}_k^{(i)}) (\vec{X} - \vec{M}_k^{(i)})^T F\{D_0^2(\vec{X}|\vec{\theta}_k^{(i)})\} \Pr(k|\vec{X}, \vec{\theta}^{(i)}) w(\vec{X}|\vec{\theta}^{(i)}) d\vec{X}}{\int \Pr(k|\vec{X}, \vec{\theta}^{(i)}) w(\vec{X}|\vec{\theta}^{(i)}) d\vec{X}}. \quad (7)$$

Равенства (5)–(7) следует понимать в хронологической последовательности итерационного взаимодействия так, что  $p_k^{(0)}$ ,  $\vec{M}_k^{(0)}$  и  $\mathbf{B}_k^{(0)}$  – начальные приближения априорной вероятности, вектора среднего и ковариационной матрицы  $k$ -го класса ( $k = 1, 2, \dots, K$ ), а  $i$  – номер текущей итерации. Ясно, что такой алгоритм не является наилучшим в смысле скорости сходимости текущей оценки вектора  $\vec{\theta}^{(i)}$  к оптимальному значению  $\theta^*$  [13, 14]. Однако он имеет наглядный статистический смысл и удобен для анализа.

### 3. ОБУЧАЮЩАЯ ВЫБОРКА

В пространстве размерности больше двух эффективным методом оценки многомерных интегралов в формулах (5)–(7) является метод *существенной выборки* [1]. В рамках этого метода погрешность вычисления, например, интеграла в числителе формулы (5)

$$P(k|\vec{\theta}) = \int \Pr(k|\vec{X}, \vec{\theta}) w(\vec{X}, \vec{\theta}) d\vec{X}$$

сильно зависит от удачного выбора распределения  $q(\vec{X}|\vec{\theta})$  для моделирования обучающей (существенной) выборки  $\{\vec{Y}(1), \dots, \vec{Y}(M)\}$ . Такой выбор должен учитывать два взаимно противоречивых фактора. Прежде всего распределение  $q(\vec{X}|\vec{\theta})$  должно допускать реализацию эффективных компьютерных алгоритмов моделирования случайных векторов  $\vec{Y}(1), \dots, \vec{Y}(M)$ . Вместе с тем, необходимо стремиться к тому, чтобы распределение  $q(\vec{X}|\vec{\theta})$  хотя бы грубо походило на оптимальное  $q_0 \Pr(k|\vec{X}, \vec{\theta}) | w(\vec{X}|\vec{\theta})$ , где  $q_0$  – нормирующая постоянная. Лишь в этом случае можно надеяться получить оценку интеграла

$$\hat{P}(k|\vec{\theta}) = \frac{1}{M} \sum_{m=1}^M \frac{\Pr(k|\vec{Y}^{(m)}, \vec{\theta}) w(\vec{Y}^{(m)}|\vec{\theta})}{q(\vec{Y}^{(m)}|\vec{\theta})} \quad (8)$$

с приемлемо малой дисперсией. С точки зрения альтернативных форм представления исходной информации: выборочной и регулярной – рациональными являются два варианта. Первый  $q(\vec{X}|\vec{\theta}) = \phi(\vec{X})$  реализует, в частности, хорошо известный *выборочный метод локального максимума правдоподобия* и отвечающие ему алгоритмы типа *EM* и *SEM* [3, 6, 11, 13, 15]. В этом нетрудно убедиться, подставив в формулу (8) равенства  $\vec{Y}^{(m)} = \vec{X}^{(m)}$ ;  $w(\vec{X}|\vec{\theta}) = \phi(\vec{X}) = q(\vec{X}|\vec{\theta})$ . Преимущество этого варианта состоит в том, что исходная выборка  $\vec{X}^{(m)}$  выступает в роли обучающей  $\vec{Y}^{(m)}$  для оценки многомерных интегралов в формулах (5)–(7). Иными словами, из алгоритма идентификации исключается этап статистического моделирования векторов  $\vec{Y}^{(1)}, \dots, \vec{Y}^{(M)}$ .

Однако распределение  $q(\vec{X}|\vec{\theta}) = \phi(\vec{X})$  отличается от оптимального  $q_0 \Pr(k|\vec{X}, \vec{\theta}) \phi(\vec{X})$  в той мере, в которой апостериорное распределение  $\Pr(k|\vec{X}, \vec{\theta})$  уклоняется от единицы. Отсюда становится очевидным основной недостаток выборочного алгоритма, а именно сильная зависимость скорости сходимости итерационного процесса от межклассовых расстояний. При малых расстояниях апостериорные распределения классов значительно уклоняются от единицы и нуля, приближаясь к величине порядка  $K^{-1}$ . Поэтому дисперсии *выборочных оценок* многомерных интегралов в формулах (5)–(7) неприемлемо большие. Кроме того, при малом объеме  $M$  исходной выборки дисперсия *эмпирического* функционала качества оценок, в частности, правдоподобия

$$\hat{L}(\vec{\theta}) = \frac{1}{M} \sum_{m=1}^M \ln \hat{\phi}(\vec{X}^{(m)}|\vec{\theta}) \quad (9)$$

становится также неприемлемо большой. В этом случае не приходится надеяться на хорошую схо-

димность итерационного процесса, максимизирующего такой функционал.

4. МОДИФИЦИРОВАННАЯ СИСТЕМА НОРМАЛЬНЫХ УРАВНЕНИЙ

В рамках регулярной формы представления исходной информации наилучшим, в смысле эффективности моделирования обучающей выборки, является второй вариант выбора распределения  $q(\vec{X}|\vec{\theta}) = \hat{\phi}_k(\vec{X}|\vec{\theta}_k)$ . Его преимущество заключается, очевидно, в необходимости разыгрывать лишь стандартные случайные векторы. В этом случае функционалы качества оценок параметров смеси целесообразно преобразовать к виду

$$L(\vec{\theta}) = \sum_{k=1}^K p_k l(k|\vec{\theta}); \quad D(\vec{\theta}) = -\ln \sum_{k=1}^K p_k d(k|\vec{\theta});$$

$$J(\vec{\theta}) = \sum_{k=1}^K p_k j(k|\vec{\theta}),$$

где  $l(k|\vec{\theta})$ ,  $d(k|\vec{\theta})$  и  $j(k|\vec{\theta})$  – взвешенные правдоподобие, расстояние и дивергенция  $k$ -го класса соответственно

$$l(k|\vec{\theta}) = \int \{ \phi(\vec{X}) / \hat{\phi}(\vec{X}|\vec{\theta}) \} \hat{\phi}_k(\vec{X}|\vec{\theta}_k) \ln \hat{\phi}(\vec{X}|\vec{\theta}) d\vec{X}; \quad (10)$$

$$d(k|\vec{\theta}) = \int \{ \phi(\vec{X}) / \hat{\phi}(\vec{X}|\vec{\theta}) \}^{1/2} \hat{\phi}_k(\vec{X}|\vec{\theta}_k) d\vec{X}; \quad (11)$$

$$j(k|\vec{\theta}) = \int \{ \phi(\vec{X}) / \hat{\phi}(\vec{X}|\vec{\theta}) - 1 \} \times \hat{\phi}_k(\vec{X}|\vec{\theta}_k) \ln \{ \phi(\vec{X}) / \hat{\phi}(\vec{X}|\vec{\theta}) \} d\vec{X}. \quad (12)$$

Таким образом, система нормальных уравнений, с учетом введенных ранее определений апостериорных распределений и весовой функции приобретает вид

$$\begin{cases} \sum_{j=1}^K p_j R(j|\vec{\theta}) - R(k|\vec{\theta}) = 0, \\ \vec{E}(k|\vec{\theta}) - Q(k|\vec{\theta}) \vec{M}_k = 0, \\ \mathbf{S}(k|\vec{\theta}) - Q(k|\vec{\theta}) \vec{M}_k \vec{M}_k^T - \phi_0 R(k|\vec{\theta}) \mathbf{B}_k = 0 \\ (k = 1, 2, \dots, K). \end{cases}$$

Здесь  $R(k|\vec{\theta})$  и  $Q(k|\vec{\theta})$  – средние веса  $k$ -го класса;  $E(k|\vec{\theta})$  и  $S(k|\vec{\theta})$  – так называемые  $w$ -взвешенные (в

смысле Л.Д. Мешалкина) оценки вектора среднего и автокорреляционной матрицы  $k$ -го класса [16];

$$R(k|\vec{\theta}) = \int W(\vec{X}|\vec{\theta}) \hat{\phi}_k(\vec{X}|\vec{\theta}_k) d\vec{X}; \quad (13)$$

$$Q(k|\vec{\theta}) = \int W(\vec{X}|\vec{\theta}) \hat{\psi}_k(\vec{X}|\vec{\theta}_k) d\vec{X}; \quad (14)$$

$$\vec{E}(k|\vec{\theta}) = \int \vec{X} W(\vec{X}|\vec{\theta}) \hat{\psi}_k(\vec{X}|\vec{\theta}_k) d\vec{X}; \quad (15)$$

$$\mathbf{S}(k|\vec{\theta}) = \int \vec{X} \vec{X}^T W(\vec{X}|\vec{\theta}) \hat{\psi}_k(\vec{X}|\vec{\theta}_k) d\vec{X}; \quad (16)$$

где  $\hat{\psi}_k(\vec{X}|\vec{\theta}_k) = -2\phi_0 F\{D_0^2(\vec{X}|\vec{\theta}_k)\} \hat{\phi}_k(\vec{X}|\vec{\theta}_k)$  – модифицированная парциальная плотность распределения;  $\phi_0$  – постоянная нормировки, полученная из условия интегрируемости с единицей модифицированной парциальной плотности распределения;  $W(\vec{X}|\vec{\theta})$  – весовая функция: для правдоподобия –

$$W(\vec{X}|\vec{\theta}) = \frac{\phi(\vec{X})}{\hat{\phi}(\vec{X}|\vec{\theta})}; \quad (17)$$

для расстояния –

$$W(\vec{X}|\vec{\theta}) = \left\{ \frac{\phi(\vec{X})}{\hat{\phi}(\vec{X}|\vec{\theta})} \right\}^{\frac{1}{2}}; \quad (18)$$

для дивергенции –

$$W(\vec{X}|\vec{\theta}) = \frac{\phi(\vec{X})}{\hat{\phi}(\vec{X}|\vec{\theta})} - 1 + \ln \frac{\phi(\vec{X})}{\hat{\phi}(\vec{X}|\vec{\theta})}. \quad (19)$$

Таким образом, алгоритм рандомизации смеси (1), альтернативный выборочному, предусматривает последовательное выполнение следующих процедур:  $S$ -шаг – статистическое моделирование обучающих выборок  $\{\vec{Y}_k^{(1)}, \vec{Y}_k^{(2)}, \dots, \vec{Y}_k^{(M)}\}$  и  $\{\vec{Z}_k^{(1)}, \vec{Z}_k^{(2)}, \dots, \vec{Z}_k^{(M)}\}$ , представляющих собой независимые в совокупности реализации стандартных случайных векторов, имеющих распределения  $\hat{\phi}_k(\vec{X}|\vec{\theta}_k)$  и  $\hat{\psi}_k(\vec{X}|\vec{\theta}_k)$ ;  $W$ -шаг – оценивание весовых функций классов и  $w$ -взвешенных начальных моментов парциальных распределений;  $M$ -шаг – оценивание оптимального вектора параметров смеси. Эти этапы рассматриваются в следующих разделах.

В заключении данного раздела отметим, что в методическом плане  $SWM$ -алгоритм напоминает метод бутстрепа (bootstrap), широко применяемый при построении непараметрических ядерных оценок вероятностных распределений [17].

Таблица 1. Модели радиальных распределений трех типов

Функция	Гаусса	Пирсона	Стьюдента
$f(\ \vec{t}\ ^2)$	$\exp\left(-\frac{\ \vec{t}\ ^2}{2}\right)$	$\left(1 - \frac{\ \vec{t}\ ^2}{2\alpha + N}\right)^{\alpha-1}$ $\alpha > 0, \ \vec{t}\ ^2 \leq 2\alpha + N$	$\left(1 + \frac{\ \vec{t}\ ^2}{\alpha + 1 - N}\right)^{-(\alpha+1)/2}$ $\alpha > N - 1$
$F(\ \vec{t}\ ^2)$	$\frac{1}{2}$	$\frac{\alpha - 1}{2\alpha + N} \left(1 - \frac{\ \vec{t}\ ^2}{2\alpha + N}\right)^{-1}$	$\frac{\alpha + 1}{2(\alpha + 1 - N)} \left(1 + \frac{\ \vec{t}\ ^2}{\alpha + 1 - N}\right)^{-1}$
$v_{N-1}$	$2^{\frac{N}{2}-1} \Gamma\left(\frac{N}{2}\right)$	$\left[(2\alpha + N)^{\frac{N}{2}} B\left(\alpha, \frac{N}{2}\right)\right]/2$	$(\alpha + 1 - N)^{\frac{N}{2}} B\left(\frac{\alpha + 1 - N}{2}, \frac{N}{2}\right)/2$
$\phi_0$	1	$(2\alpha + N)/(2\alpha + N - 2)$	2

Примечание:  $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta)$  – бета-функция;  $\varepsilon_{N, \alpha}(\vec{t}) = S_1^{-1} v_{N-1}^{-1} f(\|\vec{t}\|^2)$  и  $\omega_{N, \alpha}(\vec{t}) = -2\phi_0 F\{\|\vec{t}\|^2\} \varepsilon_{N, \alpha}(\vec{t})$ .

5. СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Эллипсоидально-симметричные распределения (2) весьма удобны для статистического моделирования обучающих выборок. Эффективный алгоритм основан на каноническом коррелирующем преобразовании [5, 18]  $\vec{Y}_k = \mathbf{U}_k \sqrt{\Lambda_k} \vec{V} + \vec{M}_k$  стандартной случайной выборки  $\{\vec{V}^{(1)}, \vec{V}^{(2)}, \dots, \vec{V}^{(M)}\}$  из генеральной совокупности, характеризуемой радиальным распределением  $\varepsilon_{N, \alpha}(\vec{V}) = S_1^{-1} v_{N-1}^{-1} f(\|\vec{V}\|^2)$  [15]. Здесь  $\Lambda_k$  и  $\mathbf{U}_k$  – матрицы собственных значений и векторов ковариационной матрицы  $\mathbf{B}_k$   $k$ -го класса, а  $\|\vec{V}\|$  – евклидова норма вектора  $\vec{V} = (v_1, v_2, \dots, v_N)^T$ . Важное преимущество такого подхода состоит в возможности моделирования вырожденного парциального распределения. Пусть в результате текущей итерации алгоритма рандомизации смеси ковариационная матрица  $k$ -го класса становится плохо обусловленной, т.е. начиная с некоторого номера  $n > N_{0k}$  ее собственные значения  $\lambda_{nk}$ , упорядоченные в порядке убывания, становятся меньше некоторого уровня значимости  $\lambda_{0k}$ . В подобном случае алгоритм моделирования обучающей выборки приобретает вид

$$\vec{Y}_k = \sum_{n=1}^{N_{0k}} (\sqrt{\lambda_{nk}} V_n) \vec{U}_{nk}; \quad \vec{V} = (V_1, V_2, \dots, V_N);$$

$$\mathbf{U}_k = (\vec{U}_{1k}, \vec{U}_{2k}, \dots, \vec{U}_{Nk}),$$

где  $\vec{U}_{nk}$  – собственный вектор матрица  $\mathbf{B}_k$ , отвечающий собственному значению  $\lambda_{nk}$ . Таким образом, альтернативный SWM-алгоритм рандомизации смеси распределений позволяет в качестве модели структуры  $N$ -мерных данных рассматри-

вать линейные многообразия размерности, меньшей чем  $N$ .

В соответствии с введенным ранее определением модифицированного парциального распределения обучающую выборку  $\{\vec{Z}_k^{(1)}, \vec{Z}_k^{(2)}, \dots, \vec{Z}_k^{(M)}\}$  также удобно моделировать с помощью канонического коррелирующего преобразования стандартной выборки  $\{\vec{T}^{(1)}, \vec{T}^{(2)}, \dots, \vec{T}^{(M)}\}$  из генеральной совокупности, характеризуемой радиальным распределением  $\omega_{N, \alpha}(\vec{t}) = -2\phi_0 F\{\|\vec{t}\|^2\} \varepsilon_{N, \alpha}(\vec{t})$ . Ряд удобных моделей радиальных распределений представлен в табл. 1. Эти распределения широко используются в качестве типовых в теории распознавания образов [4, 19] и допускают реализацию высокоэффективных компьютерных алгоритмов моделирования стандартной обучающей выборки  $\{\vec{V}^{(1)}, \vec{V}^{(2)}, \dots, \vec{V}^{(M)}\}$  [15]. Из табл. 1 нетрудно получить полезные равенства для распределения Гаусса

$$\omega_{N, \alpha}(\vec{t}) = \varepsilon_{N, \alpha}(\vec{t});$$

для распределения Пирсона:

$$\omega_{N, \alpha}(\vec{t}) = (\sqrt{\phi_0})^{-(N-1)} \varepsilon_{N, \alpha-1}(\vec{t}/\sqrt{\phi_0})/\sqrt{\phi_0};$$

для распределения Стьюдента:

$$\omega_{N, \alpha}(\vec{t}) = (\sqrt{\sigma})^{-(N-1)} \varepsilon_{N, \alpha+2}(\vec{t}/\sqrt{\sigma})/\sqrt{\sigma},$$

$$\sigma = (\alpha + 1 - N)/(\alpha + 3 - N).$$

Иными словами, для моделирования случайного вектора  $\vec{T}$  необходимы те же алгоритмы, что и для моделирования вектора  $\vec{V}$ . При этом в зависимости от типа парциального распределения возможны лишь изменения параметра  $\alpha$  и масштаба.

Таблица 2. Типы ядерных функций

Ядро	Ker( $t^2$ )	$\mu_{N-1}$
Равномерное	1	$N^{-1}$
Треугольное	$1 -  t $	$\{N(N + 1)\}^{-1}$
Епанечникова	$1 - t^2$	$2\{N(N + 2)\}^{-1}$
Дважды взвешенное	$(1 - t^2)^2$	$8\{N(N + 2)(N + 4)\}^{-1}$
Трижды взвешенное	$(1 - t^2)^3$	$48\{N(N + 2)(N + 4)(N + 6)\}^{-1}$
Нормальное	$\exp(-t^2/2)$	$2^{\frac{N}{2}-1} \Gamma(N/2)$

Примечание: Ker( $t^2$ ) = 0 для  $|t| > 1$ .

Важно отметить, что хранение в памяти компьютера стандартных обучающих выборок позволяет свести к минимуму дополнительные вычислительные затраты, возникающие на этапе статистического моделирования в SWM-алгоритме по сравнению с выборочным EM-алгоритмом.

6. НЕПАРАМЕТРИЧЕСКОЕ ОЦЕНИВАНИЕ ВЕСОВОЙ ФУНКЦИИ

Дополнительные вычислительные затраты в SWM-алгоритме появляются также на этапе непараметрического оценивания исходной многомерной плотности распределения  $\phi(\vec{X})$ , входящей в состав формул (17)–(19) для весовой функции. В рамках регулярной формы представления информации экономичный алгоритм оценки  $\phi(\vec{X})$  основан на методе кусочно-линейной интерполяции многомерной функции, заданной на регулярном ортогональном растре [20].

Применительно к выборочной форме представления информации удобной является полупараметрическая ядерная оценка исходной плотности

$$\phi^{(M)}(\vec{X}|\vec{\theta}, \mathbf{D}) = \frac{1}{M} \sum_{m=1}^M \Phi_m(\vec{X}|\vec{\theta}, \vec{D}_m). \quad (20)$$

Для структуры данных с перекрывающимися в метрике Махаланобиса эллипсоидальными кластерами заслуживает доверия оценка, построенная на основе адаптивного фильтрационного ядра [8] в виде конечной смеси

$$\Phi_m(\vec{X}|\vec{\theta}, \vec{D}_m) = \sum_{k=1}^K \Pr(k|\vec{X}^{(m)}, \vec{\theta}) \Phi_{km}(\vec{X}|\vec{\theta}, d_{km}), \quad (21)$$

где

$$\Phi_{km}(\vec{X}|\vec{\theta}, d_{km}) = J_{km} \text{Ker}\{d_{km}^{-2} D_0^2(\vec{X}|\vec{X}^{(m)}, \mathbf{B}_k)\}. \quad (22)$$

эллипсоидально-симметричные стандартные ядра.

Их веса (фильтры) пропорциональны апостериорным вероятностям наблюдений  $\Pr(k|\vec{X}^{(m)}, \vec{\theta})$ , а ориентации и масштабы – согласованы с аналогичными параметрами  $B_k$  соответствующих классов. Здесь  $J_{km}$  – нормирующий множитель, полученный из условия интегрируемости с единицей парциального ядра  $\Phi_{km}(\vec{X}|\vec{\theta}, d_{km})$ ;  $\text{Ker}\{\|\vec{v}\|^2\}$  – радиальная функция ядра с единичной дисперсией и конечным  $(N - 1)$ -м моментом

$$\mu_{N-1} = \int_0^\infty v^{N-1} \text{Ker}(v^2) dv;$$

$$J_{km} = S_1^{-1} \mu_{N-1}^{-1} (\det \mathbf{B}_k)^{-\frac{1}{2}} d_{km}^{-N},$$

а  $d_{km}$  – расстояние до ближайшего соседа оптимального ранга в метрике Махаланобиса  $k$ -го класса для  $m$ -й выборочной точки. Типы ядерных функций, широко распространенные в практических приложениях, приведены в табл. 2. На достаточно малых объемах  $M$  исходной выборки матрицу

$$\mathbf{D} = (\vec{D}_1, \dots, \vec{D}_M) = \{d_{km}\} \\ (k = 1, \dots, K), \quad (m = 1, \dots, M)$$

локальных параметров сглаживания ядер приходится оптимизировать методом скользящей проверки (cross-validation) [17]. С целью снижения вычислительных затрат эту матрицу рационально представить в виде  $\mathbf{D} = \mathbf{H} \times \Delta \mathbf{D}$ , где элементы  $h_1, \dots, h_k$  диагональной  $K \times K$  матрицы  $\mathbf{H}$  – суть подлежащие определению ранги ближайших соседей, а элемент  $\Delta d_{km}$  прямоугольной  $K \times M$  матрицы  $\Delta \mathbf{D}$  – суть фиксированный масштаб приращения расстояний до ближайших соседей в метрике Махаланобиса  $k$ -го класса для  $m$ -й выборочной точки. В отличие от работы [21] в качестве оценки масштаба  $\Delta d_{km}$ , предлагается выбирать медиану первой разности вариационного ряда расстоя-

ния  $D_0^2(\vec{X}^{(e)} | \vec{X}^{(m)}, \mathbf{B}_k)$ ,  $(e = 1, \dots, M)$ . Такого рода прием позволяет применить эффективный в вычислительном отношении одномерный поиск ранга  $h_k$  ближайших соседей в  $k$ -м классе, оптимальный по критерию максимума логарифмической выборочной функции правдоподобия

$$\hat{L}_k(\vec{\theta}, h_k) = \frac{1}{M} \sum_{e=1}^M \ln \{ \phi_{ke}^{(M)}(\vec{X}^{(e)} | \vec{\theta}, h_k) \},$$

где скользящая ядерная оценка парциального распределения  $k$ -го класса суть:

$$\begin{aligned} \phi_{ke}^{(M)}(\vec{X} | \vec{\theta}, h_k) &= \frac{1}{M-1} \times \\ &\times \sum_{\substack{m=1 \\ m \neq e}}^M \Pr(k | \vec{X}^{(m)} | \vec{\theta}) \Phi_{km}(\vec{X} | \vec{\theta}, h_k \Delta d_{km}). \end{aligned}$$

Иными словами, для фиксированного вектора параметров смеси  $\vec{\theta}$  настраиваются оптимальные ранги ближайших соседей в метриках соответствующих классов  $h_1, \dots, h_k$ . Именно в этом смысле оценка (20)–(22) является полупараметрической [7], так как представляет собой разумное сочетание параметрических и непараметрических моделей многомерной вероятностной плотности распределения.

К сожалению, полупараметрическая адаптивная фильтрационная ядерная оценка (20)–(22), как правило, обладает значительным смещением. Простой и вместе с тем эффективный метод снижения смещения предложен в работе [22]. Его вариант, обеспечивающий приемлемо малый уровень дисперсии оценки, применительно к данному случаю имеет вид

$$\tilde{\phi}^{(M)}(\vec{X} | \vec{\theta}, \mathbf{D}) = \frac{\phi^{(M)}(\vec{X} | \vec{\theta}, \mathbf{D})}{M} \sum_{m=1}^M \frac{\Phi_m(\vec{X} | \vec{\theta}, \vec{D}_m)}{\phi^{(M)}(\vec{X}^{(m)} | \vec{\theta}, \mathbf{D})}.$$

Необходимо отметить, что эта оценка не удовлетворяет условию интегрируемости с единицей и поэтому подлежит нормировке.

Применение оценки ближайших соседей создает определенные трудности при вычислении трехмерного массива  $D_0^2(\vec{X}^{(e)} | \vec{X}^{(m)}, \mathbf{B}_k)$ , расстояний Махаланобиса между наблюдениями в метрике каждого класса. Вычислительные затраты могут быть минимизированы прежде всего с помощью стратегии поиска соседей по так называемому  $k-d$ -дереву [23] и, кроме того, с помощью канонического декоррелирующего преобразования [5]  $G_{nk}^{(m)} = \vec{U}_{nk}^T \vec{X}^{(m)} / \sqrt{\lambda_{nk}}$ ,  $(n = 1, 2, \dots, N_{0k})$  исход-

ной выборки  $\{ \vec{X}^{(1)}, \vec{X}^{(2)}, \dots, \vec{X}^{(M)} \}$  и хранения в памяти компьютера массива  $N$ -мерных случайных векторов  $\vec{G}_k^{(m)} = (G_{1k}^{(m)}, G_{2k}^{(m)}, \dots, G_{N_{0k}k}^{(m)}, 0, \dots, 0)$   $(m = 1, \dots, M)$  с компонентами, некоррелированными в метрике текущих классов  $(k = 1, \dots, K)$ . Тогда  $D_0^2(\vec{X}^{(e)} | \vec{X}^{(m)}, \mathbf{B}_k) = \| \vec{G}_k^{(e)} - \vec{G}_k^{(m)} \|^2$ .

Необходимость многократного обращения ковариационных матриц  $\mathbf{B}_j$   $(j = 1, \dots, K)$  при вычислении значений парциальных распределений  $\hat{\phi}_j(\vec{X}^{(m)} | \vec{\theta}_j)$   $(j = 1, \dots, K; m = 1, \dots, M)$  создает известные неудобства при реализации выборочных алгоритмов типа *EM* и *SEM*. В *SWM*-алгоритме задача вычисления матрицы расстояний Махаланобиса  $D_0^2(\vec{Y}_k^{(m)} | \vec{\theta}_j)$  в формулах (17)–(19) эффективно решается также с помощью канонического декоррелирующего преобразования  $V_{nj} = \vec{U}_{nj}^T (\vec{Y}_k - \vec{M}_k) / \sqrt{\lambda_{nj}}$ ,  $(n = 1, 2, \dots, N_{0j}; j \neq k)$  обучающей выборки  $\{ \vec{Y}_k^{(1)}, \vec{Y}_k^{(2)}, \dots, \vec{Y}_k^{(M)} \}$ . Здесь  $\vec{V}_{jk} = (V_{1jk}, V_{2jk}, \dots, V_{N_{0j}jk}, 0, \dots, 0)$  –  $N$ -мерный стандартный случайный вектор с некоррелированными компонентами. Откуда

$$D_0^2(\vec{Y}_k^{(m)} | \vec{\theta}_j) = \begin{cases} \| \vec{V}_{jk}^{(m)} \|^2, & j = k \\ \| \vec{V}_{jk}^{(m)} \|^2, & j \neq k. \end{cases}$$

### 7. ВЗВЕШЕННЫЕ ОЦЕНКИ

Оценки  $w$ -взвешенных начальных моментов парциальных распределений смеси нетрудно получить из выражений (10)–(16) методом существенной выборки

$$\hat{l}(k | \vec{\theta}) = \frac{1}{M} \sum_{m=1}^M \left\{ \frac{\phi(\vec{Y}_k^{(m)})}{\hat{\phi}(\vec{Y}_k^{(m)} | \vec{\theta})} \right\} \ln \hat{\phi}(\vec{Y}_k^{(m)} | \vec{\theta});$$

$$\hat{d}(k | \vec{\theta}) = \frac{1}{M} \sum_{m=1}^M \left\{ \frac{\phi(\vec{Y}_k^{(m)})}{\hat{\phi}(\vec{Y}_k^{(m)} | \vec{\theta})} \right\}^{\frac{1}{2}};$$

$$\hat{j}(k | \hat{\theta}) = \frac{1}{M} \sum_{m=1}^M \left\{ \frac{\phi(\vec{Y}_k^{(m)})}{\hat{\phi}(\vec{Y}_k^{(m)} | \hat{\theta})} - 1 \right\} \ln \frac{\phi(\vec{Y}_k^{(m)})}{\hat{\phi}(\vec{Y}_k^{(m)} | \hat{\theta})};$$

$$\hat{R}(k | \vec{\theta}) = \frac{1}{M} \sum_{m=1}^M W(\vec{Y}_k^{(m)} | \vec{\theta});$$

$$\hat{Q}(k | \vec{\theta}) = \frac{1}{M} \sum_{m=1}^M W(\vec{Z}_k^{(m)} | \vec{\theta});$$

$$\hat{E}(k|\vec{\theta}) = \frac{1}{M} \sum_{m=1}^M \hat{Z}_k^{(m)} W(\hat{Z}_k^{(m)}|\vec{\theta});$$

$$\hat{S}(k|\vec{\theta}) = \frac{1}{M} \sum_{m=1}^M \hat{Z}_k^{(m)} \{ \hat{Z}_k^{(m)} \}^T W(\hat{Z}_k^{(m)}|\vec{\theta}).$$

Важно подчеркнуть, что в отличие от выборочного алгоритма рандомизации весовая функция  $W(\vec{X}|\vec{\theta})$  практически не зависит от межклассовых расстояний. Поэтому следует ожидать приемлемую скорость сходимости SWM-алгоритма рандомизации смеси, даже когда парциальные распределения существенно перекрываются в метрике Махаланобиса.

Реализация вычислений по приведенным выше формулам с помощью рекуррентных (по обучающей выборке) алгоритмов, приводит к последовательной процедуре обучения без поощрения Я.З. Цыпкина [24, 25]:

$$\begin{aligned} \hat{l}^{(m+1)}(k|\vec{\theta}) &= \hat{l}^{(m)}(k|\vec{\theta}) + \\ &+ \beta^{(m+1)} \left\{ \frac{\Phi(\hat{Y}_k^{(m+1)})}{\hat{\Phi}(\hat{Y}_k^{(m+1)}|\vec{\theta})} \ln \hat{\Phi}(\hat{Y}_k^{(m+1)}|\vec{\theta}) - \hat{l}^{(m)}(k|\vec{\theta}) \right\}; \\ \hat{d}^{(m+1)}(k|\vec{\theta}) &= \hat{d}^{(m)}(k|\vec{\theta}) + \\ &+ \beta^{(m+1)} \left\{ \left[ \frac{\Phi(\hat{Y}_k^{(m+1)})}{\hat{\Phi}(\hat{Y}_k^{(m+1)}|\vec{\theta})} \right]^2 - \hat{d}^{(m)}(k|\vec{\theta}) \right\}; \\ \hat{j}^{(m+1)}(k|\vec{\theta}) &= \hat{j}^{(m)}(k|\vec{\theta}) + \\ &+ \beta^{(m+1)} \left\{ \left[ \frac{\Phi(\hat{Y}_k^{(m+1)})}{\hat{\Phi}(\hat{Y}_k^{(m+1)}|\vec{\theta})} - 1 \right] \ln \frac{\Phi(\hat{Y}_k^{(m+1)})}{\hat{\Phi}(\hat{Y}_k^{(m+1)}|\vec{\theta})} - \hat{j}^{(m)}(k|\vec{\theta}) \right\}; \\ \hat{R}^{(m+1)}(k|\vec{\theta}) &= \hat{R}^{(m)}(k|\vec{\theta}) + \\ &+ \beta^{(m+1)} \{ W(\hat{Y}_k^{(m+1)}|\vec{\theta}) - \hat{R}^{(m)}(k|\vec{\theta}) \}; \\ \hat{Q}^{(m+1)}(k|\vec{\theta}) &= \hat{Q}^{(m)}(k|\vec{\theta}) + \\ &+ \beta^{(m+1)} \{ W(\hat{Z}_k^{(m+1)}|\vec{\theta}) - \hat{Q}^{(m)}(k|\vec{\theta}) \}; \\ \hat{E}^{(m+1)}(k|\vec{\theta}) &= \hat{E}^{(m)}(k|\vec{\theta}) + \\ &+ \beta^{(m+1)} \{ \hat{Z}_k^{(m+1)} W(\hat{Z}_k^{(m+1)}|\vec{\theta}) - \hat{E}^{(m)}(k|\vec{\theta}) \}; \\ \hat{S}^{(m+1)}(k|\vec{\theta}) &= \hat{S}^{(m)}(k|\vec{\theta}) + \\ &+ \beta^{(m+1)} \{ \hat{Z}_k^{(m+1)} [\hat{Z}_k^{(m+1)}]^T W(\hat{Z}_k^{(m+1)}|\vec{\theta}) - \hat{S}^{(m)}(k|\vec{\theta}) \}, \end{aligned}$$

где  $m = 0, 1, 2, \dots$  – индекс по стандартной обучающей выборке;  $\beta^{(m+1)} = (m+1)^{-1}$  – коэффициент усиления алгоритма обучения. Основное преимущество последовательной процедуры заключается в возможности параллельного вычисления дисперсий оценок  $w$ -взвешенных моментов и контроля соответствующих доверительных интервалов.

Для снижения дисперсий оценок в процедуру обучения рационально ввести сглаживающие (по выборке) операторы. Пусть  $est(m)$  – одна из текущих оценок взвешенных функционалов качества  $\hat{l}^{(m)}(k|\vec{\theta})$ ,  $\hat{d}^{(m)}(k|\vec{\theta})$ ,  $\hat{j}^{(m)}(k|\vec{\theta})$  или  $w$ -взвешенных начальных моментов  $\hat{R}^{(m)}(k|\vec{\theta})$ ,  $\hat{Q}^{(m)}(k|\vec{\theta})$ ,  $\hat{E}^{(m)}(k|\vec{\theta})$ ,  $\hat{S}^{(m)}(k|\vec{\theta})$  ( $k = 1, \dots, K$ ). Тогда модифицированный алгоритм одновременного действия [24] формирует оценку  $smo(m, \tilde{m}) = avr(m) + der(m)(\tilde{m} - m) + (\tilde{M} - 1)der(m)/2$ , линейно сглаженную на интервале  $m - \tilde{M} + 1 \leq \tilde{m} \leq m$ . Здесь  $\tilde{M}$  – величина интервала усреднения;

$$avr(m) = \frac{1}{\tilde{M}} \sum_{\tilde{m}=m-\tilde{M}+1}^m est(\tilde{m})$$

– скользящее среднее оценки;

$$der(m) = \frac{6}{(\tilde{M}-1)\tilde{M}(\tilde{M}+1)} +$$

$$+ \left\{ 2 \sum_{\tilde{m}=m-\tilde{M}+1}^m (\tilde{m}-m)est(\tilde{m}) + (\tilde{M}-1) \sum_{\tilde{m}=m-\tilde{M}+1}^m est(\tilde{m}) \right\}$$

– первая производная сглаженной оценки. Обучение продолжается, если выполняется хотя бы одно из следующих условий,  $|der(m)| > \varepsilon_1$  – тест абсолютной ошибки (производной);  $|der(m)| > \frac{2\varepsilon_2}{\tilde{M}-1} |avr(m)|$  – тест относительной ошибки;  $m \leq \tilde{M}$  – тест наибольшего объема стандартной обучающей выборки. В противном случае в качестве сглаженной оценки принимается значение  $smo(m, m) = avr(m) + (\tilde{M} - 1)der(m)/2$ .

### 8. ОПТИМАЛЬНЫЕ ОЦЕНКИ

Простейшая итерационная процедура решения модифицированной системы нормальных



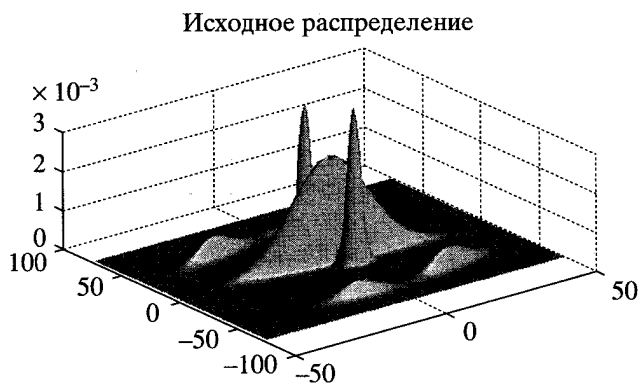


Рис. 1. Энергетический спектр изображения моря.

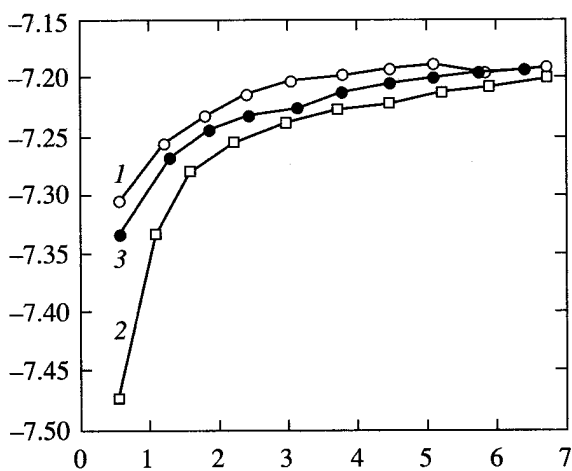


Рис. 2. Изменение правдоподобия Фишера по итерациям обучения (линейный масштаб).

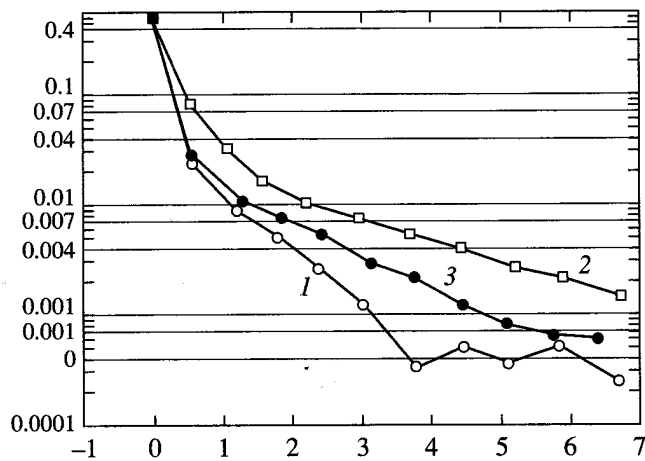


Рис. 3. Изменение расстояния Бхатачария по итерациям обучения (логарифмический масштаб); 1 – максимизация функционала правдоподобия, 2 – минимизация функционала расстояния, 3 – комбинированный функционал.

уравнений методом последовательных приближений имеет вид

$$p_k^{(i+1)} = p_k^{(i)} R(k|\vec{\theta}^{(i)}) / \sum_{j=1}^K p_j^{(i)} R(j|\vec{\theta}^{(i)})$$

$$(k = 1, \dots, K);$$

$$\vec{M}_k^{(i+1)} = \vec{E}(k|\vec{\theta}^{(i)}) / Q(k|\vec{\theta}^{(i)});$$

$$\mathbf{B}_k^{(i+1)} = \{ \mathbf{S}(k|\vec{\theta}^{(i)}) - Q(k|\vec{\theta}^{(i)}) \vec{M}_k^{(i)} [\vec{M}_k^{(i)}]^{-1} \} / \{ \phi_0 R(k|\vec{\theta}^{(i)}) \},$$

соответствующие им оценки функционалов качества параметров смеси –

$$\hat{L}(\vec{\theta}^{(i)}) = \sum_{k=1}^K p_k^{(i)} \hat{l}(k|\vec{\theta}^{(i)});$$

$$\hat{D}(\vec{\theta}^{(i)}) = -\ln \sum_{k=1}^K p_k^{(i)} \hat{d}(k|\vec{\theta}^{(i)});$$

$$\hat{J}(\vec{\theta}^{(i)}) = \sum_{k=1}^K p_k^{(i)} \hat{j}(k|\vec{\theta}^{(i)}).$$

Уточнение параметров смеси продолжается, если выполняется хотя бы одно из условий:  $|p_k^{(i+1)} - p_k^{(i)}| > \delta_1$  ( $k = 1, \dots, K$ ) – тест априорных вероятностей;  $\|\vec{M}_k^{(i+1)} - \vec{M}_k^{(i)}\| > \delta_2$ ; ( $k = 1, \dots, K$ ) – тест векторов средних;  $\|\mathbf{B}_k^{(i+1)} - \mathbf{B}_k^{(i)}\| > \delta_3$  ( $k = 1, \dots, K$ ) – тест ковариационных матриц;  $i \leq I_{\text{макс}}$  – тест наибольшего числа итераций.

### 9. ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ С SWM-АЛГОРИТМОМ

Свойства сходимости SWM-алгоритма исследовались на примере процедуры рандомизации спектральной плотности мощности (СПМ) радиолокационного изображения взволнованной морской поверхности, измеренного с помощью РЛС бокового обзора в эксперименте “Беринг” [26]. Результаты оцифровки СПМ на ортогональном растре в симметричной относительно начала координат области частотного пространства размером  $101 \times 131$  дискрет представлены на рис. 1. Начальные приближения вектора параметров  $\vec{\theta}^{\rightarrow 0}$  семи ( $K = 7$ ) спектральных классов полигауссовской модели СПМ (1) представлены в табл. 3. Значения функционалов качества обучения в точке старта процедуры рандомизации составили  $L(\vec{\theta}^{\rightarrow 0}) =$

Таблица 3. Значения начальных ( $i = 0$ ) и финальных ( $i = 10$ ) приближений параметров полигауссовской модели

Параметр	Номер класса						
	1	2	3	4	5	6	7
$p_k^{(0)}$	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429
$p_k^{(10)}$	0.0598	0.0598	0.04	0.04	0.1051	0.1051	0.5914
$M_{1k}^{(0)}$	12	-12	-7	7	3	-3	0
$M_{1k}^{(10)}$	21.99	-22.05	-10.08	9.885	4.150	-3.922	-0.783
$M_{2k}^{(0)}$	-36	36	-40	40	-8	8	0
$M_{2k}^{(10)}$	-44.15	43.91	-50.06	49.94	-10.89	10.9	-0.207
$\lambda_{1k}^{(0)}$	60	60	45	45	6	6	170
$\lambda_{1k}^{(10)}$	47.6	47.45	35.29	34.76	4.724	4.611	141.3
$\lambda_{2k}^{(0)}$	12	12	12	12	6	6	12
$\lambda_{2k}^{(10)}$	9.004	9.099	11.1	9.292	4.273	4.221	9.036
$u_{11k}^{(0)}$	1	1	1	1	1	1	1
$u_{11k}^{(10)}$	-0.894	-0.894	-0.985	-0.985	-0.821	0.5407	-0.983
$u_{12k}^{(0)}$	0	0	0	0	0	0	0
$u_{12k}^{(10)}$	-0.449	-0.448	0.1746	0.1644	-0.571	0.8412	0.1852
$u_{21k}^{(0)}$	0	0	0	0	0	0	0
$u_{21k}^{(10)}$	-0.449	-0.448	0.1746	0.1644	-0.571	-0.8412	0.1852
$u_{22k}^{(0)}$	1	1	1	1	1	1	1
$u_{22k}^{(10)}$	0.8938	0.8938	0.9846	0.9864	0.8212	0.5407	0.9827

Примечание:  $p_k$  – веса (априорные вероятности) классов,  $M_{1k}, M_{2k}$  – первая и вторая компоненты векторов средних,  $\lambda_{1k}, \lambda_{2k}$  – собственные значения ковариационных матриц,  $B_k$ , упорядоченных в порядке убывания,  $u$  – компоненты матрицы собственных векторов  $U_k$ .

$= -8.24254$ ;  $D(\vec{\theta}^0) = 0.49886$  и  $J(\vec{\theta}^0) = 5.78000$ .

Зависимости оптимальных оценок функционалов правдоподобия Фишера и расстояния Бхатачария от объема стандартной обучающей выборки

$\{\vec{V}^{(1)}, \vec{V}^{(2)}, \dots, \vec{V}^{(M)}\}$ , нормированного на величину  $M_{\text{макс}} = 6 \times 10^4$  и накопленного по итерациям  $M$ -шага процедуры рандомизации, представлены на рис. 2 и 3. Точками на графиках представлены итерации  $M$ -шага SWM-алгоритма. Численные эксперименты показали, что разумный компромисс между дисперсией оценок параметров смеси и скоростью сходимости процедуры

обучения состоит в выборе комбинированной весовой функции:

$$W(\vec{Y}|\vec{\theta}) = \begin{cases} \frac{\phi(\vec{Y})}{\hat{\phi}(\vec{Y}|\vec{\theta})} - 1 + \ln \frac{\phi(\vec{Y})}{\hat{\phi}(\vec{Y}|\vec{\theta})}, & \text{если } \frac{\phi(\vec{Y})}{\hat{\phi}(\vec{Y}|\vec{\theta})} \geq e, \\ \frac{\phi(\vec{Y})}{\hat{\phi}(\vec{Y}|\vec{\theta})}, & \text{если } 1 \leq \frac{\phi(\vec{Y})}{\hat{\phi}(\vec{Y}|\vec{\theta})} < e, \\ \left\{ \frac{\phi(\vec{Y})}{\hat{\phi}(\vec{Y}|\vec{\theta})} \right\}^{\frac{1}{2}}, & \text{если } \frac{\phi(\vec{Y})}{\hat{\phi}(\vec{Y}|\vec{\theta})} < 1, \end{cases}$$

где  $e$  – постоянная Эйлера. Такого рода взвешивание обеспечивает последовательное применение в

Модель конечной смеси распределений

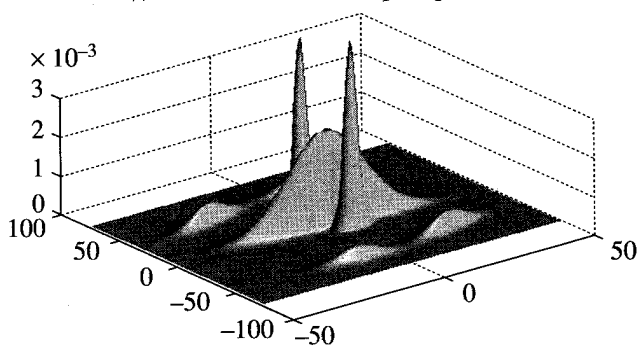


Рис. 4. Полигауссовская модель распределения, заданного случайной выборкой. 1 – максимизация функционала правдоподобия, 2 – минимизация функционала расстояния, 3 – комбинированный функционал.

Ядерная оценка распределение

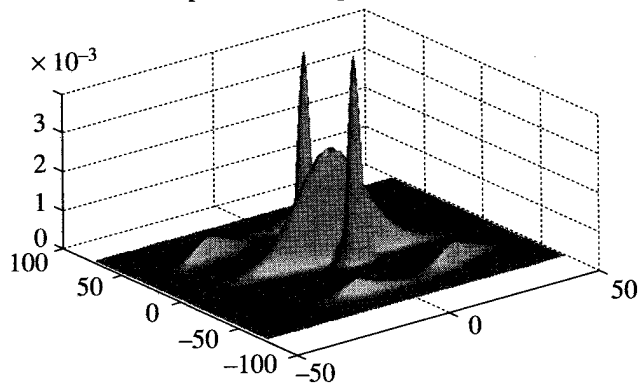


Рис. 5. Адаптивная фильтрационная ядерная оценка распределения, заданного случайной выборкой.

качестве целей обучения функционалов: дивергенции, когда ошибки значительны –  $\phi(\vec{Y})/\hat{\phi}(\vec{Y}|\vec{\theta}) \geq e$ ; правдоподобия, когда ошибки умеренные –  $1 \leq \phi(\vec{Y})/\hat{\phi}(\vec{Y}|\vec{\theta}) < e$ , и расстояния, когда ошибки незначительные, –  $\phi(\vec{Y})/\hat{\phi}(\vec{Y}|\vec{\theta}) < 1$ . Видно, что скорость сходимости процедуры рандомизации практически не хуже скорости сходимости для функционала правдоподобия. Вместе с тем дисперсии оценок параметров смеси для такой комбинированной цели обучения меньше, чем в случае правдоподобия. Десять итераций  $M$ -шага алгоритма, реализующего комбинированную цель, потребовали накопленного объема стандартной

обучающей выборки  $M_{\Sigma} = 6.4122 \times M_{\max}$ . В результате были достигнуты следующие значения оценок функционалов качества обучения:  $L(\vec{\theta}^{(10)}) = -7.19362$ ;  $D(\vec{\theta}^{(10)}) = 5.88769 \times 10^{-4}$  и  $J(\vec{\theta}^{(10)}) = 3.52432 \times 10^{-3}$ . Соответствующие оптимальные значения весов, а также характеристики положения и масштаба парциальных спектров полигауссовской модели СПМ представлены в табл. 3.

Свойства сходимости  $SWM$ -алгоритма для выборочной формы представления исходной информации исследовали на примере случайной независимой выборки объемом  $M = 1250$  реализаций двумерного вектора. Отметим, что по сравне-

Таблица 4. Невязки оценок распределения, заданного случайной выборкой, в точках парциальных средних

Параметр	Номер класса						
	1	2	3	4	5	6	7
$M_{1k}^{(10)}$	21.99	-22.05	-10.08	9.885	4.150	-3.922	-0.783
$M_{2k}^{(10)}$	-44.15	43.91	-50.06	49.94	-10.89	10.9	-0.207
$\Delta_k^{(P)}$	0.2614	0.2603	0.4316	0.4301	-3.59	-3.77	2.597
$\Delta_k^{(NP)}$	-0.119	-0.207	-0.048	-0.052	-8.66	-8.87	-1.203

Примечание:  $\Delta_k^{(P)} = \left\{ \phi(\vec{M}_k^{(10)}|\vec{\theta}^{(10)}) - \hat{\phi}(\vec{M}_k^{(10)}|\vec{\theta}^{(10)}) \right\} \times 10^4$ ;

$$\Delta_k^{(NP)} = \left\{ \phi(\vec{M}_k^{(10)}|\vec{\theta}^{(10)}) - \tilde{\phi}(\vec{M}_k^{(10)}|\vec{\theta}^{(10)}, D) \right\} \times 10^4.$$

нию с численным экспериментом, представленным в работе [27], объем выборки уменьшен в 4 раза. Выборочные значения  $(X_1^{(m)}, X_2^{(m)})$  моделировали с помощью алгоритма суперпозиции [1] полигауссовской модели плотности распределения вероятности (1) с финальными параметрами, представленными в табл. 3, и процедуры коррелирующего преобразования. Результаты работы SWM-алгоритма в этом случае иллюстрируют рис. 4 и 5, на которых соответственно представлены параметрическая оценка распределения полигауссовской моделью (1) и полупараметрическая адаптивная фильтрационная ядерная оценка (20)–(22). Абсолютные невязки этих оценок по отношению к модельному распределению (с помощью которого моделировалась исходная случайная выборка) представлены в табл. 4. Невязки вычислялись в точках математических ожиданий парциальных распределений модельной плотности вероятности.

### ЗАКЛЮЧЕНИЕ

Предложен метод рандомизации многомерных распределений с кластерной структурой в метрике Махаланобиса. Метод реализован в виде SWM-алгоритма обучения без поощрения параметров модели конечной смеси эллипсоидально симметричных распределений по экспериментальным данным. Процедура обучения адаптирована как к выборочной, так и к регулярной формам задания исходной информации. В отличие от выборочного варианта EM-алгоритма SWM-алгоритм эффективно работает с малыми выборками, малыми межклассовыми расстояниями и вырожденными распределениями.

Автор выражает благодарность профессору Б.И. Шахтарину за внимание к данной работе и ценные советы.

### СПИСОК ЛИТЕРАТУРЫ

1. Ермаков В.С. Метод Монте-Карло и смежные вопросы. М.: Наука, 1975.
2. Фомин Я.А. Теория выбросов случайных процессов. М.: Связь, 1980.
3. Duda R.O., Hart P.E. Pattern Classification and Scene Analysis. N.Y.: J. Willey and Sons, 1973.
4. Tou J.T., Gonzalez R.C. Pattern Recognition Principle. L.: Addison-Wesley P.C., 1974.
5. Fukunaga K. Introduction to Statistical Pattern Recognition. N.Y.: Acad. Press, 1972.
6. Redner R.A., Walker H.F. // SIAM Rev. 1984. V. 26. № 2. P. 195.
7. Hjort N.L., Glad I.K. // Statistical Research Report, University of Oslo, Department of Mathematics. Oslo, 1996. № 8. P. 1.
8. Marchette D.J., Priebe C.E., Rogers G.W., Solka J.L. // Comp. Statistics. 1995. V. 4. № 2. P. 27.
9. Friedman J.H. // J. Amer. Statistical Association. 1987. V. 82. № 2. P. 249.
10. Buja A., Cook D., Asimov D., Hurley C. // J. Comp. Graph. Statistics. 1999. V. 8. № 3. P. 1.
11. Hathaway R.J. // Ann. Statistics. 1985. V. 13. № 2. P. 795.
12. Wolfe J.H. // Multivariate Behavior Research. 1970. V. 5. P. 329.
13. Tanner M.A. Tools for statistical inference: Observed data and data augmentation methods. N.Y.: Springer, 1991.
14. Abaffy J., Spedicato E. ABS Projection Algorithms: Mathematical techniques for linear and nonlinear equations. N.Y.: J. Wiley and Sons, 1990.
15. Айвазян С.А., Бухштабер В.М., Енюков Н.С., Мешалкин Л.Д. Прикладная статистика: классификация и снижение размерности: Справочное издание. М.: Финансы и статистика, 1989.
16. Мешалкин Л.Д. // Уч. зап. по статистике. Прикладной многомерный статистический анализ. М.: Наука, 1978. Т. 33А. С. 11.
17. Scott D.W. Multivariate Density Estimation: Theory, Practice and Visualization. N.Y.: J. Wiley, 1992.
18. Geist J.M. // Proc. IEEE. 1979. V. 67. № 5. P. 188.
19. DeGroot M.H. Opt. Statistical Decisions. N.Y.: McGraw-Hill Co., 1970.
20. Красовский А.А. // Изв. АН СССР, Сер. Техн. кибернетика. 1989. № 3. С. 3.
21. Breiman L., Meisel W., Pursell E. // Technometrics. 1977. V. 19. № 2. P. 135.
22. Jones M.C., Linton O. // Biometrika. 1995. V. 82. № 2. P. 327.
23. Friedman J.H., Bentley J.L. // ACM Trans. math. Software. 1977. V. 3. № 3. P. 209.
24. Цыпкин Я.З. Основы теории обучающихся систем. М.: Наука, 1970.
25. Kosko B. Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence. N.Y. etc.: Prentice Hall. 1992.
26. Загородников А.А. Радиолокационная съемка морского волнения с летательных аппаратов. Л.: Гидрометеоздат, 1978.
27. Hwang J.N., Lay J.N., Lipman A. // IEEE Trans. 1994. V. Sp42. № 10. P. 2795.