ЕВРАЗИЙСКОЕ НАУЧНОЕ ОБЪЕДИНЕНИЕ
EURASIAN SCIENTIFIC ASSOCIATION
www.esa-conference.ru

О задании начальных условий для EM-алгоритма

Апраушева Н. Н., кандидат физ.-мат. наук
Сорокин С. В., программист
Вычислительный центр им. А. А. Дородницына РАН, Москва
Дикусар В. В., доктор физ.-мат. наук
Московский физико-технический институт, г. Долгопрудный, Московской области.

Введение

Модели гауссовых смесей (ГС) широко используются в качестве гибкого аппроксиматора в различных областях науки и практики [1-8]. Популярность ГС обязана таким их свойствам, как: гладкость, идентифицируемость [9], разрешимость[10], полнота в пространстве L_2 [11].

Для вычисления параметров ГС по данной выборке применяются различные алгоритмы [8], наиболее эффективным из них является EM-алгоритм (алгоритм Дэя-Шлезингера), основанный на методе максимального правдоподобия [12, 13]. Существенным недостатком этого алгоритма является трудность задания таких начальных условий, из которых последовательность итераций сходилась бы к оптимальной оценке параметра смеси. В ряде работ утверждалось [13, 14, 15], что при случайном задании начальных условий вероятность P получения оптимального решения по EM-алгоритму резко уменьшается с увеличением размерности пространства p: P=0.076 для p=5, P=0.01 для p=10, P=0.001 для p=15.

Однако, экспериментальным путём было установлено [16, 17], что

1) вероятность P является убывающей функцией параметров p, k, ε и возрастающей функцией параметров n, ρ_{is}, $i<s$, $i,s \in \{1,2,\ldots,k\}$, p — размерность пространства, k — число компонент смеси, ε — задаваемая точность вычислений ($0<\varepsilon<1$), n — объём выборки, ρ_{is} — расстояние Махаланобиса между i-й и s-й компонентами смеси;

2) при любых значениях p, k, n ($2 \le k < \infty$) существуют такие значения $\rho_0 = \rho(\rho,k,n)$, $n_0 = n(p,k,\rho_{is})$, что при всех $\rho_{is} \ge \rho_0$, $n \ge n_0$ ($0<\varepsilon<10^{-8}$) вероятность $P \ge 0.5$.

В этой работе представлены методы задания начальных условий, значительно повышающие вероятность P получения оптимальной оценки параметров смеси по EM-алгоритму. При $k=2$ дана приближённая формула для вычисления вероятности P [17]. При $k \ge 3$ метод задания начальных условий базируется на предварительной частичной классификации выборки и вычислении начального значения параметра смеси [18].

Отметим, что при малых расстояниях Махаланобиса ($\rho_{is} \le 3$) для выборок малых объёмов (n<50) EM-алгоритм даёт несколько решений, и не всегда из них оптимальным решением является то, в котором значение функции правдоподобия максимально [19, 20]. Скорректированное правило выбора оптимального решения дано в [19, 8].

1. Описание EM-алгоритма

Предполагается, что данная p-мерная выборка

$$X^{(n)} = \{X_1, X_2, \ldots, X_n\}, \tag{1.1a}$$

$$X_j = (x_{j1}, x_{j2}, \ldots, x_{jp}),\ 1 \le p < \infty, \quad j = 1,2,\ldots,n, \tag{1.1b}$$

представляет собой реализацию некоторой непрерывной p-мерной случайной величины $\eta = (\eta_1, \eta_2, \ldots, \eta_p)$ с плотностью вероятности $f(X,\theta)$, аппроксимируемой с допустимой точностью плотностью вероятности конечной гауссовой смеси с известным числом её компонент k,

$$f(X,\theta) \approx \frac{(2\pi)^{-\frac{p}{2}}}{|\Sigma|^{\frac{1}{2}}} \sum_{i=1}^{k} \pi_i e^{-\frac{1}{2}(X-\mu_i)\Sigma^{-1}(X-\mu_i)'},\ 2 \le k < \infty, , \tag{1.2}$$

π_i — априорная вероятность i-й компоненты, μ_i — вектор её среднего значения, Σ — ковариационная матрица каждой компоненты,

$$\theta = (\mu_1, \mu_2, \ldots, \mu_k, \pi_1, \pi_2, \ldots, \pi_{k-1}, \Sigma),\ 2 \le k < \infty, \quad \pi_i \ge 0, \tag{1.3a}$$

$$\pi_i \ge 0,\ \sum_{i=1}^{k} \pi_i = 1, \tag{1.3 b}$$

Для удобства неизвестный векторный параметр θ представим в виде

ЕВРАЗИЙСКОЕ НАУЧНОЕ ОБЪЕДИНЕНИЕ
EURASIAN SCIENTIFIC ASSOCIATION
www.esa-conference.ru

$$\theta = (\theta_1, \theta_2, \ldots, \theta_m). \tag{1.4}$$

Число компонент m параметра θ равно

$$m = 2^{-1} p(p+1) + k(p+1) - 1. \tag{1.5}$$

Для вычисления оптимальной оценки параметр θ по выборке (1.1a) используем EM-алгоритм, основанный на методе максимального правдоподобия. При довольно слабых условиях регулярности функции плотности вероятности среди множества решений системы уравнений правдоподобия $\{\tilde{\theta}_l\}$, $l = 1, 2, \ldots, u$, существует только одно оптимальное решение $\tilde{\theta}_{opt}$ — состоятельное, асимптотически эффективное и асимптотически нормальное, и в нём функция правдоподобия имеет локальный максимум. Гауссова смесь с равными ковариационными матрицами удовлетворяет условиям регулярности, сформулированным в работах [21, 22]. Логарифмическая функция правдоподобия $\ln L\left(X^{(n)}, \theta\right)$ гауссовой смеси (1.2) имеет вид:

$$\ln L\left(X^{(n)}, \theta\right) = -\frac{pn}{2}\ln(2\pi) - \frac{n}{2}\ln|\Sigma| + \sum_{j=1}^{n}\ln\left[\sum_{i=1}^{k}\pi_i e^{-\frac{1}{2}(X_j-\mu_i)\Sigma^{-1}(X_j-\mu_i)'}\right], \tag{1.6}$$

система уравнений правдоподобия определяется равенствами:

$$\frac{\partial \ln L\left(X^{(n)}, \theta\right)}{\partial \theta_i} = 0,\ i = 1, 2, \ldots, s. \tag{1.7}$$

Следуя Крамеру [23], мы называем *оценкой максимального правдоподобия любое решение системы уравнений правдоподобия.* Система уравнений (1.7) при использовании выражений апостериорных вероятностей для каждой точки выборки X_j,

$$P_{ij} = \frac{\pi_i e^{-\frac{1}{2}(X_j-\mu_i)\Sigma^{-1}(X_j-\mu_i)'}}{\sum_{s=1}^{k}\pi_s e^{-\frac{1}{2}(X_j-\mu_i)\Sigma^{-1}(X_j-\mu_i)'}},\ i = 1, 2, \ldots, k,\ j = 1, 2, \ldots, n, \tag{1.8}$$

приводится к следующему виду [13]:

$$\pi_i = \frac{1}{n}\sum_{j=1}^{n} P_{ij}, \tag{1.9}$$

$$\mu_i = \frac{\sum_{j=1}^{n} X_j P_{ij}}{\sum_{j=1}^{n} P_{ij}}, \tag{1.10}$$

$$\Sigma = \sum_{j=1}^{n}\sum_{i=1}^{k}(X_j - \mu_i)'(X_j - \mu_i) P_{ij}. \tag{1.11}$$

Если ввести обозначения

$$a_i = \Sigma^{-1}\mu_i' + \tau,\ i = 1, 2, \ldots, k, \tag{1.12a}$$

$$b_i = -\frac{1}{2}\mu_i\Sigma^{-1}\mu_i' + \ln\pi_i + \nu,\ i = 1, 2, \ldots, k, \tag{1.12b}$$

где τ и ν — произвольные вектор и скаляр соответственно, то формулы (1.8) представимы следующим образом:

$$P_{ij} = 1 / \sum_{s=2}^{k} e^{(a_s - a_i)X_j' + b_s - b_i}, \tag{1.13}$$

$$i = 1, 2, \ldots, k-1,\ j = 1, 2, \ldots, n.$$

ЕВРАЗИЙСКОЕ НАУЧНОЕ ОБЪЕДИНЕНИЕ
EURASIAN SCIENTIFIC ASSOCIATION
www.esa-conference.ru

Из выражений (1.13) следует, что параметры π_i, μ_i, $i=1,2,\ldots,k$, Σ, в (1.9)-(1.11) являются функциями a_i, b_i, $i=1,2,\ldots,k$, X_j, $j=1,2,\ldots,n$. Тогда уравнения (1.12a), (1.12b) образуют ряд уравнений общего вида:

$$a_i=\phi_i\left(a_i^{(t-1)},b_i^{(t-1)};X_1,X_2,\ldots,X_n\right),\ i=1,2,\ldots,k, \tag{1.14a}$$

$$b_i=\psi_i\left(a_i^{(t-1)},b_i^{(t-1)};X_1,X_2,\ldots,X_n\right),\ i=1,2,\ldots,k,. \tag{1.14b}$$

где $a_i^{(t)}$, $b_i^{(t)}$ — значения параметров a_i, b_i на t-м шаге итеративной процедуры. По выбранным начальным значениям $a_i^{(0)}$, $b_i^{(0)}$, $i=1,2,\ldots,k$, и формулам (1.13) и (1.9)-(1.11) соответственно определяем значения апостериорных вероятностей $P_{ij}^{(0)}$ и параметров смеси ($\pi_i^{(0)}$; $\mu_i^{(0)}$; $\Sigma_i^{(0)}$; $i=1,2,\ldots,k$). Далее, полученные значения $\{\pi_i^{(0)};\ \mu_i^{(0)};\ \Sigma_i^{(0)};\ i=1,2,\ldots,k\}$ используются для вычисления $P_{ij}^{(1)}$, $i=1,2,\ldots,k$, $j=1,2,\ldots,n$, по формулам (1.12a), (1.12b), (1.13) и т. д. Доказано [13, 16], что последовательность

$$\theta^{(t)}=\left(\pi_i^{(t-1)},\mu_i^{(t-1)},\Sigma_i^{(t-1)},\quad i=1,2,\ldots,k\right),\ t=1,2,\ldots, \tag{1.15}$$

сходится к решению системы уравнений правдоподобия, которое является точкой локального максимума или седловой точкой функции правдоподобия.

При различных начальных условиях a_{0i}, b_{0i}, $i=1,2,\ldots,k$, для одной выборки могут получаться различные решения системы уравнений правдоподобия. Начальные значения векторов a_{0i} задаём случайными, а параметры b_{0i} находим из равенств

$$b_{0i}=-\left(a_{0i},\tilde{X}_{(i)}\right), \tag{1.16}$$

полагая, что начальные гиперплоскости, разделяющие компоненты смеси

$$a_iX'+b_i=0,\ i=1,2,\ldots,k. \tag{1.17}$$

проходят либо через общий центр тяжести — точку

$$\tilde{X}=\frac{1}{n}\sum_{j=1}^{n}X_j, \tag{1.18}$$

и тогда в (1.15) имеем: $\tilde{X}_{(i)}=\tilde{X}$, $i=1,2,\ldots,k$, либо через k точек исходной выборки, взятых случайным образом, и тогда

$$\tilde{X}_{(i)}=X_j,\ j\in\{1,2,\ldots,n\},\ i=1,2,\ldots,k. \tag{1.19}$$

2. О задании начальных условий для k=2

Для k=2 выражения (1.13) представимы в виде:

$$P_{1j}=\frac{1}{1+\exp\left(aX_j'+b\right)},\ P_{2j}=1-P_{1j},\ j=1,2,\ldots,n, \tag{2.1}$$

где

$$a'=\Sigma^{-1}\left(\mu_2-\mu_1\right)',\ b=\frac{1}{2}\left(\mu_1\Sigma^{-1}\mu_1'-\mu_2\Sigma^{-1}\mu_2'\right)+\ln\left(\frac{\pi_2}{\pi_1}\right). \tag{2.2}$$

Для k=2 уравнение гиперплоскости, разделяющей компоненты смеси (линейной дискриминантной функции Фишера) имеет вид [23]:

$$X\Sigma^{-1}\left(\mu_2-\mu_1\right)'+\frac{1}{2}\left(\mu_1\Sigma^{-1}\mu_1'-\mu_2\Sigma^{-1}\mu_2'\right)+\ln\left(\frac{\pi_2}{\pi_1}\right)=0. \tag{2.3}$$

Из выражений в (2.2), (2.3) следует, что вектор a' есть нормаль к гиперплоскости в уравнении (2.3), а скаляр b — свободный член этого уравнения. При проведении экспериментов в двумерном пространстве было обнаружено, что при случайном задании множества начальных векторов a_{i_0} $i=1,2,\ldots,50$, существует

ЕВРАЗИЙСКОЕ НАУЧНОЕ ОБЪЕДИНЕНИЕ
EURASIAN SCIENTIFIC ASSOCIATION
www.esa-conference.ru

аварийный вектор a_a, лежащий в окрестности вектора a_{ort}, ортогонального прямой, проходящей через точки μ_1 и μ_2. Эта прямая делит исследуемую выборку на две группы, не являющиеся классами. На рис. 1 изображена эта ситуация, $\tilde{\mu}_1$, $\tilde{\mu}_2$ — оценки средних значений этих групп, аварийная окрестность заштрихована. Скалярное произведение $\left(a_{\text{ort}}, \mu_2 - \mu_1\right) = 0$. Для каждой выборки существует множество аварийных начальных векторов, лежащих в угловой окрестности вектора a_{ort} [16, 17].

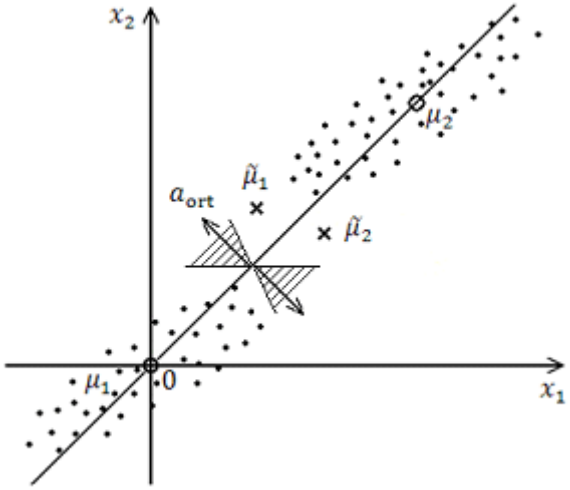

Рис. 1.

Аварийный вектор a_0 можно обнаружить на t-ой итерации, если для скалярного произведения векторов выполняется условия:

$$\left|\left(\mu_2^{(t)} - \mu_1^{(t)}\right), \Sigma^{-1(t)}\left(\mu_2^{(t)} - \mu_1^{(t)}\right)\right| < \varepsilon, \tag{2.4}$$

ε — малое положительное число, $\varepsilon < 10^{-2}$. Тогда итерационный процесс прерывается и выбирается новый начальный вектор a_0.

На основании многочисленных экспериментов для различных значений p, ρ, n ($\varepsilon = 10^{-8}$) на модельных выборках были вычислены статистические оценки $\tilde{P}_{\text{ст}}$ вероятности P и подобранна приближённая формула для вычисления этой вероятности [17],

$$\tilde{P}_f = \left(1 - \frac{\sqrt{p}\,p}{\rho\sqrt{n}}\right)\left(1 - \frac{p}{\rho\sqrt{n}}\right), \ \rho \geq 2. \tag{2.5}$$

По данным таблицы в [17] значения $\tilde{P}_{\text{ст}}$ и $\tilde{P}_f$ отличаются незначительно. Если $\Delta P = \left|\tilde{P}_{\text{ст}} - \tilde{P}_f\right|$, то для n=400 имеем [17]: $\max \Delta P = 0.14$, $\min \Delta P = 0.02$, $\overline{\Delta P} = 0.06$.

Так как $P\left(p, n, \rho\right)$ — возрастающая функция параметра ρ, то для фиксированных значений p, n, ε имеем нижнюю границу для $\tilde{P}_f$, $\tilde{P}_f\left(p, n, \rho\right) > \tilde{P}_f\left(p, n, 2\right)$, в правой части неравенства $\rho = 2$.

3. О задании начальных условий для $k \geq 3$

В этом случае использование вышеизложенного метода задания начальных условий может привести к резкому уменьшению значений вероятности P при увеличении числа компонент смеси k и размерности выборки p. Кроме того, добавляется ещё один неблагоприятный фактор — взаимное расположение гауссианов. Поэтому целесообразно для задания начальных условий провести частичную классификацию элементов выборки $X^{(n)}$ и вычисление начальной оценки векторного параметра $\tilde{\theta}_0$ в (1.3а) [18]. Для использования терминов кластер-анализа дадим определение кластера (класса).

Определение 3.1. Кластер множества $X^{(n)}$ — это его непустое подмножество ω_s, $s \in \{1, 2, \ldots, k\}$, $k \geq 1$, описываемое с допустимой точностью некоторой унимодальной функцией плотности вероятности (гауссианом в нашем случае).

ЕВРАЗИЙСКОЕ НАУЧНОЕ ОБЪЕДИНЕНИЕ
EURASIAN SCIENTIFIC ASSOCIATION
www.esa-conference.ru

Тогда исходная выборка $X^{(n)}$ представима в виде

$$X^{(n)} = \bigcup_s \omega_s \,,\ s = 1,2,\ldots,k \,,\ k \geq 1 \,. \tag{3.1}$$

Для проведения предварительного анализа структуры множества $X^{(n)}$ зададим на нём подходящую метрику, например евклидову, и вычислим расстояния между всеми его элементами по формуле:

$$r_{ij} = \left(\sum_{m=1}^{p} \left(x_{im} - x_{jm} \right)^2 \right)^{\frac{1}{2}} ,\ i < j \,,\ i = 1,2,\ldots,n-1 \,,\ j = 2,3,\ldots,n \,. \tag{3.2}$$

Упорядочив по возрастанию множество $\left\{ r_{ij} \right\}$, получим вариационный ряд,

$$\text{ВР: } r_{(1)} \leq r_{(2)} \leq \ldots \leq r_{(s)} \,,\ s = n\left(n-1\right)/2 \,. \tag{3.3}$$

Строится гистограмма ВР, аппроксимирующая его неизвестную плотность вероятности $\varphi\left(r\right)$, рис. 2.

Утверждение 3.1. Если расстояния между всеми кластерами (гауссианами) велики, то гистограмма ВР имеет хотя бы один устойчивый локальный минимум (ЛМИН).

На рис. 2 гистограмма ВР имеет ЛМИН на отрезке $\left[r_q, r_{q+1} \right]$ и два локальных максимума на отрезках $\left[r_v, r_{v+1} \right]$ и $\left[r_u, r_{u+1} \right]$.

Определение 3.1. ЛМИН, наблюдаемый на отрезке $\left[r_q, r_{q+1} \right]$ рис. 2 гистограммы, назовём **устойчивым, статистически значимым** (СЗЛМИН), если на выбранном уровне значимости α в минимальном промежутке $\left[r_q, r_{u+1} \right]$, содержащим этот ЛМИН и наименьший из двух ближайших к нему ЛМАКС, отвергается гипотеза H_0 о постоянстве функции $\varphi\left(r\right)$,

$$H_0 :\ \varphi\left(r\right) = \left(r_{u+1} - r_q \right)^{-1} ,\ r \in \left[r_q, r_{u+1} \right] .$$

Для обнаружения статистически значимых локальных минимумов гистограмм были использованы статистические критерии Колмогорова, ω^2, χ^2, Вилкоксона (корректно используемый для зависимых случайных величин $r_{(i)}$ [18, 24, 25]).

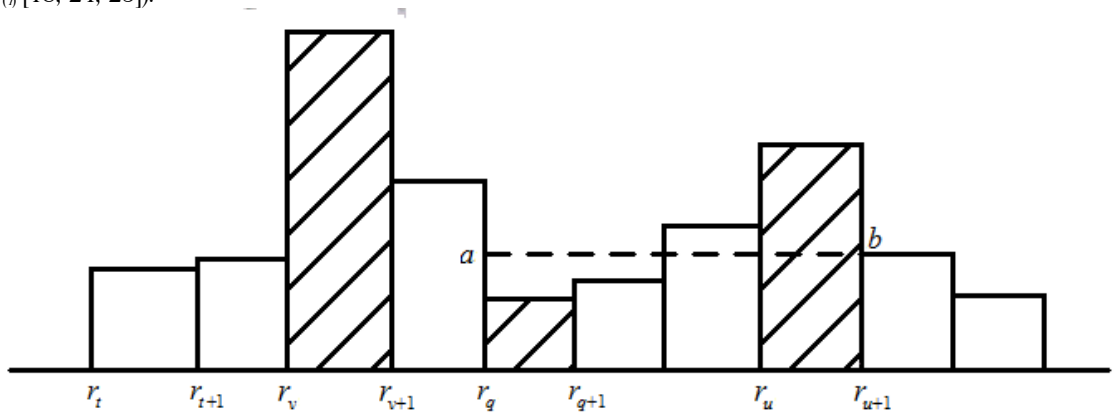

Рис. 2.

Возможна ситуация, когда неизвестная функция $\varphi\left(r\right)$ в некотором интервале имеет локальный минимум, но он не обнаруживается на её гистограмме используемыми статистическими критериями.

Если гистограмма имеет несколько статистически значимых локальных минимумов, то фиксируется первый из них (нумерация идёт слева направо), который даёт оценку наибольшего диаметра кластера $d_{\max}$,

$$d_{\max} = r_q \,. \tag{3.4}$$

Из данной выборки $X^{(n)}$ методом случайного выбора без возвращения формируем репрезентативную подвыборку L объёма n_1 [26],

ЕВРАЗИЙСКОЕ НАУЧНОЕ ОБЪЕДИНЕНИЕ
EURASIAN SCIENTIFIC ASSOCIATION
www.esa-conference.ru

$$L = \left\{ X'_1, X'_2, \ldots, X'_{n_1} \right\},\ k < n_1 < n,\ n\pi_{\min} \geq 2, \tag{3.5}$$

И строится её вариационный ряд.

Далее проводится классификация элементов множества L методом ближайшего соседа при пороговом значении $\mathrm{r}_{\text{пор}} = d_{\max}$ [27]. При завершении процесса классификации для всех точек множества L имеем начальные оценки для параметров смеси.

$$\tilde{\mu}_{i_0} = \left(\sum_{s=1}^{n_{i_0}} X'_s \right) / n_{i_0},\ \tilde{\pi}_{i_0} = n_{i_0} / n_1,\ \tilde{\Sigma}_0 = \tilde{\Sigma}_{(\max)}, \tag{3.6}$$

n_{i_0} — число точек множества L, попавших в класс ω_{i_0}, $\tilde{\Sigma}_{(\max)}$ — ковариационная матрица класса $\omega_{i_0} \subset L$, содержащего наибольшее число элементов.

Если все кластеры множества $X^{(n)}$ близко расположены друг к другу, то гистограмма функции $\varphi(r)$ имеет один локальный максимум или статистически незначимые локальные минимумы. В этом случае пороговое значение $\mathrm{r}_{\text{пор}}$ подбирается: $\mathrm{r}_{\text{пор}} = \mathrm{r}_q$, r_q — правый конец отрезка, содержащего ЛМАКС (рис. 2), $\mathrm{r}_{\text{пор}} = \left(\mathrm{r}_{(s1)} - \mathrm{r}_{(11)} \right) / \nu$, $\mathrm{r}_{(11)}$ — первый член ВР множества L, $\mathrm{r}_{(s1)}$ — его последний член, $\nu = 2,3,\ldots$, значение ν подбирается, оптимальным считается такое ν_0, при котором число классов $\tilde{k}$ равно заданному значению k.

Литература:

1. Titterington D. M., Smith A. F. M., and Makov U. E. Statistical Analysis of Finite Mixture Distributions. JornWiley, 1985. Pp. 53-147.

2. Carreira-Perpicбn N. A. Mode-finding for mixture of Gaussian distributions. // IEEE Trans. On Pattern Analys. and Mach. Intell. — V. 22, № 11. — 2000. — P. 1318-1323.

3. Фомин Я. А., Тарловский Г. Р. Статистическая теория распознавания образов. — М.: Радио и связь, 1986. — 264 с.

4. Harris N. and Smith S. A. B. The sib-sib age of on set correlation among individuals suffering from a hereditary syndrome produced by more than one gene. // Annals of Eugenics. London, 1949. Vol. 14. Part 4. Pp. 309-318.

5. Di Crescenzo A., Martinucci B. On a symmetric nonlinear birth-death process with bimodal transition probabilities // Symmetry, 2009. Vol. 1. N. 2. Pp.201-214.

6. Aprausheva N. N., Gorlach I. A., Zhelnin A. A., Sorokin S. V. An experiment on Automated Statistical Recognition of Clouds. // J. Computational Mathematics and Mathematical Physics, 1998. Vol. 38. № 10. Pp. 1715-1719.

7. Reynolds D. A. Speaker identification and verification using Gaussian mixture speaker modls. Speech Communication. 1995. Vol. 17. Pp. 91-108.

8. Апраушева Н. Н., Сорокин С. В. Заметки о гауссовых смесях. — М.: Вычислительный центр им. А. А. Дородницына РАН, 2015.— 144 с.

9. Yakowitz S. J., Spragins J. D. On the identificability of finite mixtures. Ann. Math. Stat., 1968. 39. N. 1. Pp. 209-214.

10. Хартиган Дж. А. Задачи, связанные с функциями распределения в кластер-анализе // Классификация и кластер / (Перевод с англ.). М.: Мир, 1980. С. 42-65.

11. Дороднов А. А. Ортонормированная система Гаусса. Сборник аспирантских работ. Точные науки. КГУ, Казань, 1969.

12. Шлезингер М. И. Взаимосвязь обучения и самообучения в распознавании образов. // Киев: Кибернетика. — 1968.— № 2. — С. 81-88.

13. Day N. E. Estimating the Components of a Mixture of Normal Distributions // Biometrika. — 1969. — V. 56, № 3. — P. 463-477.

14. Айвазян С. А., Бежаева З. И., Староверов О. В. Классификация многомерных наблюдений. — М.: Статистика, 1974. — 239 с.

15. Бежаева З. И. О содержании библиотеки программ по разделу «Классификация и снижение размерности». Алгоритмическое и программное обеспечение прикладного статистического анализа. — М.: Наука, 1980. — С. 163-188.

16. Апраушева Н. Н. Об использовании смеси нормальныйх распределений в распознавании образов. Диссертация. М.: ВЦ АН СССР, 1981. — 136 с.

17. Апраушева Н. Н., Сорокин С. В. О вероятности нахождения оптимальной оценки параметра двухкомпонентной гауссовой смеси, получаемой по ЕМ-алгоритму. Сб. Статистические методы оценивания и проверки гипотез. Пермский гос. университет. Вып. 27, 2016. — С. 1-11.

18. Апраушева Н. Н. Новый подход к обнаружению кластеров. М.: Вычислительный центр РАН,1993,65 с.

19. Апраушева Н. Н. Об оптимальном решении системы уравнений правдоподобия смеси анормальных распределений. Деп. ВИНИТИ 1979, № 371579.

20. Hawkins R. N. A Note on Multiple Solutions to the Mixed Distributions Problem . Tehnometrics, 1972. Vol. 114. N. 4.

21. Kulldorff G. On the Condition for Consisttency and Asymptotic Efficiency of Maximum Lilelihood Estimates // Skzndinavisk Aktuaritidskrift. — 1957. — V. 40. P. 129-144.

22. Norden R. H. A Survey of Maximum Likelihood Estimation: Part 1 // International Statistical Review. — 1972. — V. 40, № 3. — P. 329-354/

23. Крамер Г. Математические методы статистики. М.: Мир, 1975, 648 с.

24. Большев Л. Н., Смирнов Н. В. Таблицы математической статистики. М.: Наука, 1983. 416 с.

25. Орлов А. И. Некоторые вероятностные вопросы теории классификации. Прикладная статистика. Учёные записки по статистике. М.: Наука, 1983, 350 с.

26. Феллер В. Введение в теорию вероятностей и её приложения. М.: Мир, 1984. Т. 1. 527 с.

27. Загоруйко Н. Г., Елкина В. Н., Лбов Г. С. Алгоритмы обнаружения закономерностей. Новосибирск: Наука СО, 1985. 110 с.