ОЦІНЮВАННЯ ПАРАМЕТРІВ ҐАВСІВСЬКОЇ СУМІШІ РОЗПОДІЛІВ ЗА МЕТОДОМ ЕМ.

Калюжний О. Я., доктор фіз.-мат. наук, професор
КПІ ім. Ігоря Сікорського, м. Київ, Україна

Популярними в наш час статистичними моделями випадкових сигналів є скінчені декомпозиції розподілів ймовірностей вигляду [1]:

$$p\left(\mathbf{X}/\boldsymbol{\alpha},\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\ldots,\boldsymbol{\theta}_K\right)=\sum_{q=1}^{K}\alpha_q\cdot f_q\left(\mathbf{X}/\boldsymbol{\theta}_q\right), \quad (1)$$

$\boldsymbol{\alpha}=\left(\alpha_1,\alpha_2,\ldots,\alpha_K\right)$ — вектор вагових коефіцієнтів, на які накладаються обмеження:

$$\alpha_q\geq 0\,,\ \sum_{q=1}^{K}\alpha_q=1\ . \quad (2)$$

$f_q\left(\mathbf{X}/\boldsymbol{\theta}_q\right)$ — багатовимірна густина розподілу ймовірностей спостережень $\mathbf{X}$ q-го класу, $\boldsymbol{\theta}_q$ — відповідний вектор параметрів. Завдання полягає в оцінюванні по вибірці спостережень сигналу, що підлягає розподілу (1), вектору параметрів $\boldsymbol{\Theta}=\left[\boldsymbol{\alpha},\ \boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\ldots,\boldsymbol{\theta}_K\right]$.

Неважко бачити, що з врахуванням обмежень (2) декомпозиція (1) є середнім розподілом спостережень, одержаним із K неоднорідних статистичних ансамблів. При цьому вагові коефіцієнти $\alpha_1,\ldots,\alpha_K$ відповідають за змістом апріорним ймовірностям приналежності спостережень до відповідних ансамблів. Практичне застосування таких моделей зумовлене двома групами причин. По-перше, декомпозиція (1) може використовуватися як наближення складних розподілів через суперпозицію якихось простіших розподілів. По-друге, моделі у вигляді суміші розподілів мають місце у задачах, де масив спостережень $\mathbf{X}$ й фактично є статистично неоднорідним. Подібні ситуації виникають, наприклад, у задачах розпізнавання та класифікації нестаціонарних сигналів, зокрема, мовних. З математичної точки зору оцінювання параметрів декомпозиції (1) є досить складною задачею. Метою доповіді є наочна ілюстрація вирішення цієї задачі за допомогою відомого [1, 2] алгоритму ЕМ (від Expectation-Maximization).

Припустимо, як це часто вважають, що масив спостережень по кожному ансамблю $\mathbf{X}_q$ складається з N векторів $\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N$, кожен з яких відповідає окремому статистичному випробуванню, поміж собою ці випробування статистично незалежні. Тоді для спільного розподілу спостережень q-го класу на підставі теореми множення ймовірностей можна записати:

$$f_q\left(\mathbf{X}/\boldsymbol{\theta}_q\right)=\prod_{i=1}^{N}f_q\left(\mathbf{x}_i/\boldsymbol{\theta}_q\right),$$

де $f_q\left(\mathbf{x}/\boldsymbol{\theta}_q\right)$ — спільний розподіл компонент кожного з векторів $\mathbf{x}_i$. Запишемо логарифм функції правдоподібності задачі:

$$\log p\left(\mathbf{X}/\boldsymbol{\Theta}\right) = \sum_{i=1}^{N} \log\left(\sum_{q=1}^{K} \alpha_q \cdot f_q\left(\mathbf{x}_i/\boldsymbol{\theta}_q\right)\right). \quad (3)$$

Безпосередня максимізація цього функціоналу по інформаційних параметрах $\boldsymbol{\Theta}$ ускладнюється тим, що під знаком логарифму міститься сума розподілів. Однак, цю складність можна обійти, якщо штучно ввести індикаторний вектор $\boldsymbol{y} = (y_1, y_2, \ldots, y_N)$, кожен елемент якого вказує на приналежність відповідного вектору спостережень до певного ансамблю, тобто, $y_i = q$, якщо вектор $\mathbf{x}_i$ має розподіл $f_q\left(\mathbf{x}/\boldsymbol{\theta}_q\right)$. Тоді виразом для логарифму функції правдоподібності повного вектору спостережень $\left[\mathbf{y}, \mathbf{X}\right]$ є:

$$\log p\left(\mathbf{y}, \mathbf{X}/\boldsymbol{\Theta}\right) = \sum_{i=1}^{N} \log\left(p\left(\mathbf{x}_i / y_i, \boldsymbol{\Theta}\right) \cdot p\left(y_i/\boldsymbol{\Theta}\right)\right) = \sum_{i=1}^{N} \log\left(\alpha_{y_i} \cdot f_{y_i}\left(\mathbf{x}_i/\boldsymbol{\theta}_{y_i}\right)\right) \quad (4)$$

де враховано ту обставину, що $p\left(y_i/\boldsymbol{\Theta}\right) = \alpha_{y_i}$. Тепер цей індикаторний вектор $\mathbf{y}$ будемо розглядати, як “втрачені” дані та скористаємося для пошуку оцінки інформаційних параметрів $\boldsymbol{\Theta}$ алгоритмом ЕМ. Перш за все треба знайти так звану Q-функцію задачі [1]:

$$Q\left(\boldsymbol{\Theta}, \hat{\boldsymbol{\Theta}}^{(k)}\right) = E\left\{\log\left[p\left(\mathbf{y}, \mathbf{X}/\boldsymbol{\Theta}\right)\right] / \mathbf{X}, \hat{\boldsymbol{\Theta}}^{(k)}\right\} = \sum_{\mathbf{y} \in \mathbf{Y}} \log\left[p\left(\mathbf{y}, \mathbf{X}/\boldsymbol{\Theta}\right)\right] \cdot p\left(\mathbf{y}/\mathbf{X}, \hat{\boldsymbol{\Theta}}^{(k)}\right),$$

де $\hat{\boldsymbol{\Theta}}^{(k)}$ — поточне наближення оцінки. З використанням (4), знаходимо:

$$Q\left(\boldsymbol{\Theta}, \hat{\boldsymbol{\Theta}}^{(k)}\right) = \sum_{q=1}^{K} \sum_{i=1}^{N} \log\left(\alpha_q\right) \cdot W_i^{(k)}(q) + \sum_{q=1}^{K} \sum_{i=1}^{N} \log\left(f_q\left(\mathbf{x}_i/\boldsymbol{\theta}_q\right)\right) \cdot W_i^{(k)}(q),$$

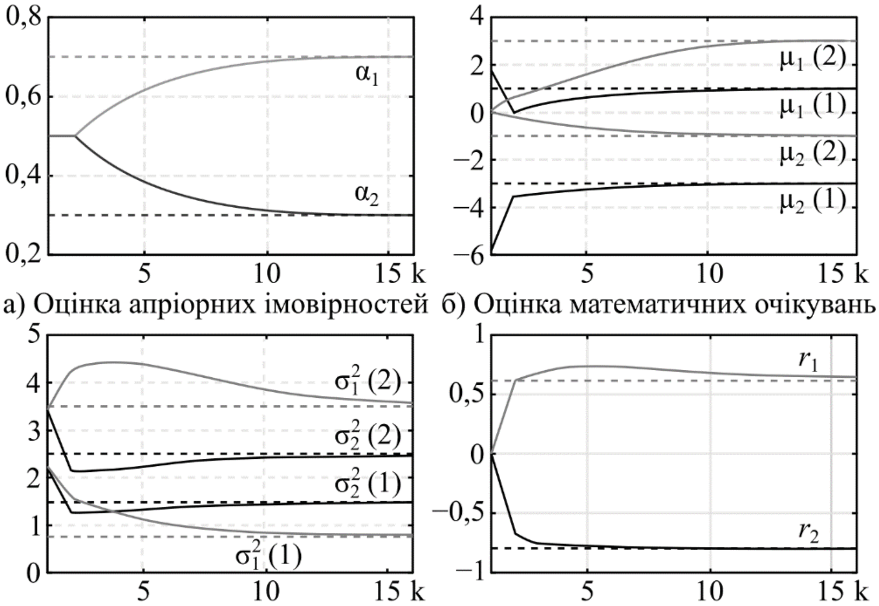
де $W_i^{(k)}(q) = p\left(y_i = q/\mathbf{x}_i,\ \hat{\boldsymbol{\Theta}}^{(k)}\right)$. Наступна ітерація оцінки $\hat{\boldsymbol{\Theta}}^{(k+1)}$ може бути одержана шляхом максимізації цієї функції по $\boldsymbol{\Theta}$. Зокрема, максимізація першого додатку цього виразу методом невизначених коефіцієнтів Лагранжа з урахуванням обмежень (2) дозволяє знайти наступну ітерацію оцінки вагових коефіцієнтів $\hat{\alpha}_q^{(k+1)}$. Апостеріорний розподіл $W_i^{(k)}(q)$ індикаторних даних знаходиться за формулою Байєса. Таким чином:

$$\hat{\alpha}_q^{(k+1)} = \frac{1}{N} \sum_{i=1}^{N} W_i^{(k)}(q),\ W_i^{(k)}(q) = \frac{\hat{\alpha}_q^{(k)} \cdot f_q\left(\mathbf{x}_i/\hat{\boldsymbol{\theta}}_q^{(k)}\right)}{\sum_{l=1}^{K} \hat{\alpha}_l^{(k)} \cdot f_l\left(\mathbf{x}_i/\hat{\boldsymbol{\theta}}_l^{(k)}\right)}.$$

Тепер розглянемо чисельний приклад. Нехай кожний елемент спостережень $\mathbf{X}$ є двовимірним вектором (x_1, x_2), що може належати одному з двох можливих ґавсівських розподілів. Спостереження цих двох класів об'єднуються до загального масиву $\mathbf{X}$ шляхом чергування у випадковому порядку.

Відносна кількість спостережень різних класів у загальній вибірці визначається відповідними апріорними ймовірностями α_1 та α_2. Мета статистичних випробувань полягає в оцінюванні математичних очікувань, дисперсій та коефіцієнтів кореляції компонентів векторів (x_1, x_2) кожного з двох класів.

При статистичному моделюванні було вибрано такі параметри двох класів ґавсівських розподілів: математичні очікування $\boldsymbol{\mu}_1 = (1,\ 3)^T$, $\boldsymbol{\mu}_2 = (-3,\ -1)^T$, дисперсії $\boldsymbol{\sigma}_1^2 = (0{,}8,\ 3{,}5)^T$, $\boldsymbol{\sigma}_2^2 = (1{,}5,\ 2{,}5)^T$, коефіцієнти кореляції $r_1 = 0{,}6$, $r_2 = -0{,}8$, апріорні ймовірності класів $\alpha_1 = 0{,}3$, $\alpha_2 = 0{,}7$.



а) Оцінка апріорних імовірностей б) Оцінка математичних очікувань

в) Оцінка дисперсій г) Оцінка коефіцієнтів кореляції

Рисунок 1. Ілюстрація збіжності оцінок параметрів класів за алгоритмом ЕМ

Рис. 1 ілюструє збіжність оцінок параметрів двох розподілів. Неважко бачити, що оцінки досягають сталих значень вже через 10 кроків від початку процесу. Але треба зазначити, що успіх суттєво залежить від початкової точки старту.

Перелік посилань

1. McLachlan G. The EM Algorithm and Extensions / G. McLachlan, T. Krishnan — N.Y.: John Wiley & Sons, Inc. 1997. — 288 p. ISBN 0-471-12358-7.

2. F. Gu, H. Zhang, W. Wang, and S. Wang, “An expectation-maximization algorithm for blind separation of noisy mixtures using Gaussian mixture model,” Circuits, Systems, and Signal Processing, vol. 36, no. 7, pp. 2697–2726, 2017

Анотація

Розглянуто чисельний приклад вирішення задачі розділення ґавсівської суміші розподілів за алгоритмом ЕМ. Цей приклад може бути корисним для студентів, аспірантів, молодих науковців при вивченні математичних основ статистичного оброблення сигналів.

Ключові слова: Expectation-Maximization, статистичне оброблення сигналів.

Abstract

A numerical example for problem of Gaussian mixture separation by means of EM algorithm is given. The example can be useful for students, postgraduates, young researchers under studying mathematical bases of statistical signals processing.

Keywords: Expectation-Maximization, statistical processing of signals.