On the set of optimal initial conditions in the Dai-Schlesinger's algorithm

Aprausheva H.H., Dikusar B.B. and Sorokin B.S.

FIZ correctional RAN, Russia
plat@ccas.ru,dikussar@yandex.ru,www2013@ccas.ru

Dai-Schlesinger algorithm often called EM-algorithm in the scientific literature, for calculating the unknown parameters of the Gaussian mixture available sample. To determine the parameters of the mixture are used well-known classical methods, modification of these methods and implement their programs. To calculate the optimal parameter estimates multidimensional blend the most effective and commonly used algorithm is the Dai-Schlesinger, based on two methods: maximum likelihood and Picard successive approximations. Use in various fields of science and practice of the Gaussian mixture model with equal covariance matrices and vectors with different mean values due to its resistance to violations of the assumptions of normality and completeness of the system of Gaussian functions in space $L_2(-\infty, \infty)$.

A number of studies with reference to the source noted that the probability P optimal solutions for algorithm Dai-Schlesinger at random the initial conditions dramatically decreases with increasing dimension p of the sample space. However, experimentally is found that the probability P is a decreasing function parameter p, k, ε, (k – the number of components of the mixture, ε – accuracy of calculations), and an increasing function of the parameter ρ_{is}, n (ρ_{is} – Mahalanobis distance between the components of the mixture, $i < s$, $i, s \in \{1, 2, \ldots, k\}$, n – sample size). It should be approved on the basis of analysis of experiments.

When $k \geq 2$ any dimension p of the sample space there exist $\rho_0 = \rho(k, p)$, $n_0 = n(k, p)$, ε_0 $(0 < \varepsilon_0 < 10^{-8})$ that, for all $\rho_{is} \geq \rho_0$, $n \geq n_0$, $0 < \varepsilon < \varepsilon_0$ likely to get the optimal solution $P \geq 0,5$.

In addition, in this study it adjusted the rule for choosing the optimal solution that takes into account the frequency of each solution at a random initial conditions.