Precise EM-based LCDG-approximation of a Probability Distribution

I. Sequential initialization of the LCDG model

The initial LCDG model closely approximating a given marginal gray level distribution $\mathbf{F}$ is built using the conventional EM-algorithm [1]–[3] adapted to the DGs. The approximation involves the following steps:

1) The distribution $\mathbf{F}$ is approximated with a mixture $\mathbf{P}_K$ of K positive DGs, relating each to a dominant mode.
2) Deviations between $\mathbf{F}$ and $\mathbf{P}_K$ are approximated with the scaled sign-alternate "subordinate" DGs as follows:
    a) The positive and negative deviations are separated and scaled up to form two temporay "probability distributions" $\mathbf{D}^{\mathrm{p}}$ and $\mathbf{D}^{\mathrm{n}}$.
    b) The same conventional EM algorithm is used iteratively to find a subordinate mixture of positive or negative DGs that approximates best $\mathbf{D}^{\mathrm{p}}$ or $\mathbf{D}^{\mathrm{n}}$, respectively (the sizes $C_{\mathrm{p}} - K$ and C_{n} of these mixtures are found by minimizing sequentially the total absolute error between each "distribution" $\mathbf{D}^{\mathrm{p}}$ or $\mathbf{D}^{\mathrm{n}}$ and its mixture model by the number of the components).
    c) The obtained positive and negative subordinate mixtures are scaled down and then added to the dominant mixture yielding the initial LCDG model of the size $C = C_{\mathrm{p}} + C_{\mathrm{n}}$.

The resulting initial LCDG has K dominant non-negative weights, $w_{\mathrm{p},1}, \ldots, w_{\mathrm{p},K}$, such that $\sum_{r=1}^{K} w_{\mathrm{p},r} = 1$, and a number of subordinate smaller non-negatgive weights such that $\sum_{r=K+1}^{C_{\mathrm{p}}} w_{\mathrm{p},r} - \sum_{l=1}^{C_{\mathrm{n}}} w_{\mathrm{n},l} = 0$.

II. Final refinement of the LCDG model

The initial LCDG is refined by approaching the closest local maximum of the log-likelihood

$$L(\mathbf{w},\mathbf{\Theta}) = \sum_{q\in\mathbf{Q}} f(q)\log p_{\mathbf{w},\mathbf{\Theta}}(q) \tag{1}$$

with the EM process adapting that in [4] to the DGs. The latter process extends in turn the conventional EM algorithm in [2] onto the sign-alternate linear combinations.

Let $p^{[m]}_{\mathbf{w},\mathbf{\Theta}}(q) = \sum_{r=1}^{C_{\mathrm{p}}} w^{[m]}_{\mathrm{p},r}\psi(q|\theta^{[m]}_{\mathrm{p},r}) - \sum_{l=1}^{C_{\mathrm{n}}} w^{[m]}_{\mathrm{n},l}\psi(q|\theta^{[m]}_{\mathrm{n},l})$ denote the current LCDG at iteration m. Relative contributions of each signal $q \in \mathbf{Q}$ to each positive and negative DG at iteration m are specified by the respective conditional weights

$$\pi^{[m]}_{\mathrm{p}}(r|q) = \frac{w^{[m]}_{\mathrm{p},r}\psi(q|\theta^{[m]}_{\mathrm{p},r})}{p^{[m]}_{\mathbf{w},\mathbf{\Theta}}(q)};\quad \pi^{[m]}_{\mathrm{n}}(l|q) = \frac{w^{[m]}_{\mathrm{n},l}\psi(q|\theta^{[m]}_{\mathrm{n},l})}{p^{[m]}_{\mathbf{w},\mathbf{\Theta}}(q)} \tag{2}$$

such that the following constraints hold:

$$\sum_{r=1}^{C_{\mathrm{p}}} \pi^{[m]}_{\mathrm{p}}(r|q) - \sum_{l=1}^{C_{\mathrm{n}}} \pi^{[m]}_{\mathrm{n}}(l|q) = 1;\ q = 0,\ldots,Q-1 \tag{3}$$

The following two steps iterate until the log-likelihood is increasing and its changes become small:

- **E– step**$^{[m]}$: Find the weights of Eq. (2) under the fixed parameters $\mathbf{w}^{[m-1]}$, $\mathbf{\Theta}^{[m-1]}$ from the previous iteration $m-1$, and
- **M– step**$^{[m]}$: Find the conditional MLEs $\mathbf{w}^{[m]}$, $\mathbf{\Theta}^{[m]}$ by maximizing $L(\mathbf{w},\mathbf{\Theta})$ under the fixed weights of Eq. (2).

Considerations closely similar to those in [1]–[3] show this process converges to a local log-likelihood maximum. As shown in a very general way in [3] and for the LCDG model in [4], it is actually a block relaxation MM-process. Let the log-likelihood of Eq. (1) be rewritten in the equivalent form with the constraints of Eq. (3) as unit factors:

$$\begin{aligned} L(\mathbf{w}^{[m]},\mathbf{\Theta}^{[m]}) \;=\; & \sum_{q=0}^{Q} f(q)\left[\sum_{r=1}^{C_{\mathrm{p}}} \pi^{[m]}_{\mathrm{p}}(r|q)\log p^{[m]}(q)\right] \\ -\; & \sum_{q=0}^{Q} f(q)\left[\sum_{l=1}^{C_{\mathrm{n}}} \pi^{[m]}_{\mathrm{n}}(l|q)\log p^{[m]}(q)\right] \end{aligned} \tag{4}$$

Let the terms $\log p^{[m]}(q)$ in the first and second brackets be replaced with the equal terms $\log w^{[m]}_{\mathrm{p},r} + \log\psi(q|\theta^{[m]}_{\mathrm{p},r}) - \log\pi^{[m]}_{\mathrm{p}}(r|q)$ and $\log w^{[m]}_{\mathrm{n},l} + \log\psi(q|\theta^{[m]}_{\mathrm{n},l}) - \log\pi^{[m]}_{\mathrm{n}}(l|q)$, respectively, which follow from Eq. (2).

At the E-step, the constrained Lagrange maximization of the log-likelihood of Eq. (4) under the Q constraints of Eq. (3) results just in the weights $\pi^{[m+1]}_{\mathrm{p}}(r|q)$ and $\pi^{[m+1]}_{\mathrm{n}}(l|q)$ of Eq. (2) for all $r = 1,\ldots,C_{\mathrm{p}}$; $l = 1,\ldots,C_{\mathrm{n}}$ and $q \in \mathbf{Q}$. At the M-step, the DG weights $w^{[m+1]}_{\mathrm{p},r} = \sum_{q\in\mathbf{Q}} f(q)\pi^{[m+1]}_{\mathrm{p}}(r|q)$ and $w^{[m+1]}_{\mathrm{n},l} = \sum_{q\in\mathbf{Q}} f(q)\pi^{[m+1]}_{\mathrm{n}}(l|q)$ follow from the constrained Lagrange maximization of the log-likelihood in Eq. (4) under the constraints $\sum_{r=1}^{C_{\mathrm{p}}} w_{\mathrm{p},r} - \sum_{l=1}^{C_{\mathrm{n}}} w_{\mathrm{n},l} = 1$ and the fixed conditional weights of Eq. (2). Under these constraints, the conventional MLEs of the parameters of each DG stem from maximizing the log-likelihood after each difference of the cumulative Gaussians is replaced with its close Gaussian density approximation (below "c" stands for "p" or "n", respectively):

$$\mu^{[m+1]}_{\mathrm{c},r} = \frac{1}{w^{[m+1]}_{\mathrm{c},r}} \sum_{q\in\mathbf{Q}} q\cdot f(q)\pi^{[m+1]}_{\mathrm{c}}(r|q)$$

$$(\sigma^{[m+1]}_{\mathrm{c},r})^2 = \frac{1}{w^{[m+1]}_{\mathrm{c},r}} \sum_{q\in\mathbf{Q}} \left(q - \mu^{[m+1]}_{\mathrm{c},i}\right)^2 \cdot f(q)\pi^{[m+1]}_{\mathrm{c}}(r|q)$$

This modified EM-algorithm is valid until the weights $\mathbf{w}$ are strictly positive. The iterations should be terminated when the log-likelihood of Eq. (1) almost does not change or begins to decrease due to accumulation of rounding errors.

The final mixed LCDG-model $p_C(q)$ is partitioned into the K LCDG-submodels $\mathbf{P}_{[k]} = [p(q|k) : q \in \mathbf{Q}]$, one per dominant mode, or class $k = 1,\ldots,K$, by associating the subordinate DGs with the dominant terms so that the misclassification rate is minimal.

REFERENCES

[1] A. Webb, *Statistical Pattern Recognition*, J. Wiley & Sons, New York, 2002.

[2] M. I. Schlesinger and V. Hlavac, *Ten Lectures on Statistical and Structural Pattern Recognition*, Kluwer Academic, Dordrecht, 2002.

[3] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, New York, 2001.

[4] G. Gimel'farb, A. A. Farag, and A. El-Baz, "Expectation-Maximization for a linear combination of Gaussians," *Proc. 17th IAPR Int. Conf. Pattern Recognition*, Cambridge, UK, August 23–26, 2004, IEEE CS Press, vol. 3, pp. 422–425.