Correction to: "A **Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Lawrence R. Rabiner, Proc. IEEE, Feb. 1989**

p. 271, **Comparison of HMMs (based on correction provided by Vladimir Petroff)**

There is an error in the equation before numbered equation 88; the correct equation should read:

$$s = \frac{p+q-2pq-r}{1-2r}$$

The term $-r$ was inadvertently omitted from the equation in the tutorial.

p. 272, **Scaling Section** (After equation (90)).

Consider the computation of $\alpha_t(i)$. We use the notation $\alpha_t(i)$ to denote the unscaled α's, $\hat{\alpha}_t(i)$ to denote the scaled (and iterated) α's and $\hat{\hat{\alpha}}_t(i)$ to denote the local version of α before scaling. For each t, we first compute $\hat{\hat{\alpha}}_t(i)$ according to the induction formula (20), in terms of the previously scaled $\hat{\alpha}_t(i)$ i.e.,

$$\hat{\hat{\alpha}}_t(i) = \sum_{j=1}^{N} \hat{\alpha}_{t-1}(j)\, a_{ji} b_i(O_t) \tag{91a}$$

We determine the scaling coefficient c_t as

$$c_t = \frac{1}{\sum_{i=1}^{N} \hat{\hat{\alpha}}_t(i)} \tag{91b}$$

giving

$$\hat{\alpha}_t(i) = c_t\, \hat{\hat{\alpha}}_t(i) \tag{91c}$$

so that equation (91a) can be written as

$$\hat{\hat{\alpha}}_t(i) = \sum_{j=1}^{N} c_t\, \hat{\hat{\alpha}}_{t-1}(j)\, a_{ji}\, b_i(O_t) \tag{92a}$$

We can now write the scaled $\hat{\alpha}_t(i)$ as

$$\hat{\alpha}_t(i) = \frac{\sum_{j=1}^{N} \hat{\hat{\alpha}}_{t-1}(j)\, a_{ji}\, b_i(O_t)}{\sum_{i=1}^{N}\sum_{j=1}^{N} \hat{\hat{\alpha}}_{t-1}(j)\, a_{ji}\, b_i(O_t)} \tag{92b}$$

By induction we can write $\hat{\alpha}_{t-1}(j)$ as

$$\hat{\alpha}_{t-1}(j) = \left(\prod_{\tau=1}^{t-1} c_\tau \right) \alpha_{t-1}(j) \tag{93a}$$

Thus we can write $\hat{\alpha}_t(i)$ as

$$\begin{aligned} \hat{\alpha}_t(i) &= \frac{\sum_{j=1}^{N} \alpha_{t-1}(j) \left(\prod_{\tau=1}^{t-1} c_\tau \right) a_{ji}\, b_i(O_t)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_{t-1}(j) \left(\prod_{\tau=1}^{t-1} c_\tau \right) a_{ji}\, b_i(O_t)} \\ &= \frac{\alpha_t(i)}{\sum_{i=1}^{N} \alpha_t(i)} \end{aligned} \tag{93b}$$

i.e., each $\alpha_t(i)$ is effectively scaled by the sum over all states of $\alpha_t(i)$.