Outline of the Dynamic Topology Representing Network

The DTRN was proposed by Si, Lin and Vuong (2000). A simplified explanation of its working principles is elaborated here. Like its non-dynamic counterpart, the TRN, the DTRN encodes a data manifold **X** with probability distribution *P* (x) into a finite set of reference vectors ('prototypes', centroids; typical brand profiles in this application) while respecting the topological properties of the observed data. The quantization techniques which are topologically sensitive are characterized by monitoring the neighbourhood structure of their prototypes. This information is stored in an adjacency matrix with zero/one entries and gets updated in each training iteration. Unlike the popular K-means cluster procedure the neighbourhood structure in the (D)TRN permits indirect updates of the centroids. In analogy to the fuzzy K-means or overlapping K-centroids clustering (Chaturvedi, Carroll, Green and Rotondo 1997) this increases the robustness of the quantization results.

The similarity between a data point and a prototype is measured by the Euclidean distance *d* between the *i*-th prototype's co-ordinates ('weights') vector $\mathbf{w}_i$ and an input data vector **x** with values $x_1, \ldots, x_V$

$$d_i = \|\mathbf{x} - \mathbf{w}_i\| = \left(\sum_{v=1}^{V} (x_v - w_{iv})^2\right)^{\frac{1}{2}} \tag{A-1}$$

The TRN and DTRN were inspired by the Self-Organizing Map (Kohonen, 1982) which employs stochastic approximation ('training') to adapt its weight structure according to the distribution pattern of the input data. Each of the prototypes thus learns to represent a homogeneous subset of data vectors. In the DTRN the number of such prototypes is not predetermined as the training starts with just one prototype equal to an input vector randomly selected from the data set **X**. Another randomly chosen data point **x** is compared to this first prototype i=1 according to (A-1). If d_i fails to drop below the vigilance threshold ρ, the **x** becomes a second prototype $\mathbf{w}_g$.

Once there are three or more prototypes they begin to compete with each other such that the winner i^* with

$$\|\mathbf{x} - \mathbf{w}_{i^*}\| < \|\mathbf{x} - \mathbf{w}_i\|, \quad \forall\, i, \tag{A-2}$$

and the co-winner $i^{**}$with

$$\|\mathbf{x} - \mathbf{w}_{i^{**}}\| < \|\mathbf{x} - \mathbf{w}_i\|, \quad \forall\, i \neq i^*, \tag{A-3}$$

become eligible for a weight update. Before that the winner is subject to the vigilance test. If it fails a new prototype *g* is introduced and takes the values of the current data point **x**. The adjacency matrix **S** indicating the connectivity among the prototypes is then updated in the following manner:

$$s_{gj} = \begin{cases} 1 & if\ j = i^* \\ 0 & else \end{cases} \tag{A-4}$$

$$t_{gj} = \begin{cases} 0 & if\ j = i^* \\ \infty & else \end{cases} \tag{A-5}$$

where:

$t_{g\,j}$ is an age counter denoting the number of iterations covered since the creation or last refreshment of the connection $s_{g\,j}$.

If the winner i^* passes the vigilance test i^* and all its neighbours get updated by the following 'winner-takes-quota' learning rule:

$$\Delta \mathbf{w}_{i^*}(k) = s_{i^*,i}\, \lambda(k) \frac{\exp(-\eta(k)\left\|\mathbf{x}(k)-\mathbf{w}_{i^*}(k)\right\|^2)}{\sum_{j=1}^{L} s_{i^*,j} \exp(-\eta(k)\left\|\mathbf{x}(k)-\mathbf{w}_{i^*}(k)\right\|^2)} (\mathbf{x}(k)-\mathbf{w}_{i^*}(k)), \quad i = 1,...,L \qquad \text{(A-6)}$$

where:

$0 < \lambda(k) < 1$ is the learning rate that decays with the growing number of iterations
$k = 0,1,...;$
$\eta(k)$ is an annealing factor that increases during the training.

The last two steps in the DTRN procedure regard updating the connection lifetime record and the removal of superfluous prototypes. Age correction occurs via $t_{i^*,j} = t_{i^*,j} + 1$ and the removal of outdated connections, i.e. setting $s_{i^*,\,j} = 0$, happens for an age counter exceeding the lifetime limit, i.e. $t_{i^*,j} > \tau$. A prototype i becomes redundant and is abolished if all its connections s_{ij} are zero.

The crucial parameter is the vigilance factor which controls the dynamic creation and demolition of prototypes. Si et al. suggest a schedule such as

$$\rho = \rho_0 \left(\frac{\rho_1}{\rho_0}\right)^{k/k_{\max}} \quad \text{with } \rho_0 > \rho_1 \qquad \text{(A-7)}$$

and a maximum number of k_{max} iterations; this makes ρ gradually decrease from ρ_0 to ρ_1. The authors also provide ample evidence of the DTRN performance on synthetic data with known properties and thereby offer advice on choosing meaningful parameter settings.

References

Chaturvedi, Anil, J. Douglas Carroll, Paul E. Green and John A. Rotondo (1997). A Feature-Based Approach to Market Segmentation Via Overlapping K-centroids Clustering. *Journal of Marketing Research* 34 (August), 370-77.

Kohonen, Teuvo (1982). Self-organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics*, 43, 59-69.

Si, J., S. Lin, and M.-A. Vuong (2000). Dynamic Topology Representing Networks. *Neural Networks*, 13, 617-27.

J. A. Mazanec / 05-2004