Questions and Answers

Chapter 6

Questions

(Q1) *What is the structure of AFLC?*

(Q2) *Does AFLC have to initialize like k-means? if not, why?*

(Q3) *How does AFLC dynamically adjust the number of clusters?*

(Q4) *What is the difference between DA and SA?*

(Q5) *What is the DA cost function and what does it minimize?*

(Q6) *What effect does the temperature reduction rate parameter have on DA clustering?*

(Q7) *How does DA adjust the number of clusters?*

(Q8) *What does mass-constrained DA mean?*

(Q9) *What makes MS segmentation different from normal brain segmentation?*

(Q10) *Judging from the examples given in the chapter, what are the performance differences among AFLC, DA, FCM and k-means?*

(Q11) *What is the limitation of clustering segmentation based on image intensity?*

(Q12) *How is clustering in retinal optic disk/cup and blood vessel segmentation better than regular edge detection techniques?*

(Q13) *Why is registration necessary in 3-D retinal disc/cup segmentation and how is it done?*

(Q14) *How is the 3-D optic disc/cup map created?*

Answers

(A1) AFLC has a two-layer structure. The first layer is a self-organizing neural network similar to ART1. It employs a MAXNET learning rule to classify incoming samples into corresponding clusters. The second layer serves as a verification process that performs a vigilance test. Depending whether a sample passes the test, the system is optimized or new cluster is created.

(A2) No. The total number of clusters produced by AFLC is controlled by a user-provided parameter. Then the final number of clusters depends on the distribution of the income data. Therefore, the total number of clusters is not known as in contrast to k-means, for which the total number of clusters is known before clustering starts. AFLC initializes by taking the first sample as the first centroid.

(A3) The number of clusters produced by AFLC is controlled by a user-provided parameter. When an incoming sample is initially classified to a class, it is checked if it truly belongs to the class. This is done by computing a ratio between the distance of the sample to the winning class centroid and the average distance of all the sample sin this cluster to the centroid. If the ratio is higher than the user-provided threshold, the sample is excluded from that class, a new class is created with this sample as the initial centroid.

(A4) Simulated annealing is an optimization algorithm based on Metropolis algorithm that mimics the annealing process in order to obtain global minimum. However, SA moves randomly on the energy surface and converges to a configuration of minimal energy very slowly, if the control parameter T is lowered no faster than logarithmically. DA improves the speed of convergence; the effective energy is deterministically optimized at successively reduced T while maintaining the annealing process aiming at global minimum.

(A5) DA minimizes the expected distortion of the given system ensemble while maximizing its randomness (Shannon's entropy). The cost function is expressed as

$$F = D - TH,$$

where D is the average distortion given by

$$D = \sum_x \sum_y p(x, y)d(x, y) = \sum_x p(x) \sum_y p(y \mid x)d(x, y),$$

and H the system entropy

$$H(X, Y) = -\sum \sum p(x, y) \log p(x, y).$$

(A6) The temperature reduction rate is crucial in trying to reach global minimum of the system. Analogous to the annealing process in statistical mechanics, when the temperature is reduced sufficiently slowly, the system reaches absolute minimum energy state; otherwise, a suboptimal energy state will be reached. To achieve global minimum, the temperature reduction rate has to be sufficiently large such that the system is "cool" slowly.

(A7) DA increases the number of clusters in the system through cluster splitting. As the temperature is lowered, it gradually approaches the critical temperature of the cluster. When it reaches the critical temperature, the splitting occurs. The critical temperature of a cluster is given by $T_c = 2\lambda_{\max}$, where $\lambda_{\max}$ is the maximum eigenvalue of the covariance matrix of the posterior distribution $p(x \mid y)$ of the cluster corresponding to centroid y:

$$C_{x|y} = \sum_x p(x \mid y)(x - y)(x - y)^t.$$

(A8) It can happen that the cluster annealing process has a certain dependence on the number of centroids in each effective cluster, which has to do with the perturbation during the initial partition. To solve this problem, mass-constrained DA, in which the condition $\sum_i p_i = 1$ is applied. Here p_i is the centroids that coincide in the same cluster i at position y_i.

(A9) In normal brain segmentation, usually the gray matter, white matter and CSF are to be segmented. These tissues are more distinguishable with better contrast in pixel intensity levels such that they can be separated by clustering. In MS segmentation, however, the lesions areas possess similar pixel intensities to those tissues nearby. Such lack of contract often results in misclassifying MS lesions into one of the GM, WM and CSF category, making it difficult to separate MS lesions from them.

(A10) Both k-means and FCM suffer from the initialization and local minimum problems. Cluster initialization is crucial in yielding satisfactory results. When not initialized properly, a clustering algorithm might be trapped in a local minimum, failing to proceed to the correct cluster. AFLC is an automated and adaptive improvement over k-means and FCM by incorporating neural leader clustering and FCM. The performance is improved; however, the similar problems are still encountered. Initialization is eliminated by selecting the first incoming sample as initial centroid, therefore, the outcome is sample-order dependent. DA claims to be able to achieve global minimum with faster "cooling" than classical simulated annealing. It does not have initialization problem. It is not sensitive to parameter tuning, and is noise tolerant and guaranteed to converge. The segmentation examples shown in this work, for instance, the MR image segmentation, demonstrates that DA performs better than the other three algorithms.

(A11) Clustering segmentation based solely on image pixel intensity can be affected by noise. Besides, since isolated pixel value is the only considered feature, it might fail to exploit other information in the image such as the relationships among neighborhood pixels and thus miss shapes and edges, affecting the accuracy of segmentation.

(A12) Challenge of blood vessel segmentation lies in distinguishing the blood vessel edges. However, the retinal images are usually noisy and non-uniformly illuminated. In the optic disk area, blood vessel edges are smeared by reflectance. The efficiency of simple segmentation techniques such as edge detection is reduced since they can be easily affected by noisy and incapable to distinguish edges between different tissues. For example, edges of the optic disk can be mistakenly classified as blood vessel edges. DA clustering solves this problem because of the classification is based on the intensities of the pixels, which tell the difference among them. When combined with regular edge detection technique and morphological filtering, the effect of noise can be removed, and accurate blood vessel edges can be found. Image resulted directly from DA segmentation is still affected by the noise in the original image, since single pixel intensity is used as feature. After morphological filtering, the noise in the segmented image can be easily removed. The expansion or shrinking of blood vessel edges caused by morphological erosion and dilation can be corrected by edges detected by a simple edge detector, such as a Canny edge detector.

(A13) The stereo disc image pairs must be registered in order to find the disparity between the images accurately. The disparity map allows the creation of depth maps for the optic cup/disc and subsequent 3-D segmentation of disc/cup from contour sliced at different depths.

(A14) 3-D optic disc/cup maps are created by assuming a nonconvergent stereo imaging system allowing a simple relationship between the disparity and depth for corresponding pixels in the stereo pairs while the accuracy of the disparity map is dependent on a number of preprocessing steps including feature extraction and registration of the stereo pairs combining power cepstrum and zero-mean-normalized-cross-correlation techniques which extracts depth information using coarse-to-fine disparities between corresponding windows in a stereo pair.