CT Image Classification by Threshold Circuits

A. Albrecht[1], E. Hein[2], D. Melzer[2], K. Steinhöfel[3], and M. Taupitz[2]

[1] Dept. of Computer Science,
Univ. of Hertfordshire, Hatfield, Herts AL10 9AB, UK
Email: A.Albrecht@herts.ac.uk
[2] Faculty of Medicine, Inst. of Radiology,
Humboldt University of Berlin, 10117 Berlin, Germany
Email: {eike.hein,daniela.melzer,matthias.taupitz}@charite.de
[3] GMD–National Research Center for Information Technology,
Kekuléstraße 7, 12489 Berlin, Germany
Email: Kathleen.Steinhoefel@gmd.de

Abstract. We present an algorithm that computes a depth-three threshold circuit for the classification of liver tissue. The circuit is calculated from a sample set S of 348 positive (abnormal findings) and 348 negative (normal liver tissue) examples by a local search strategy. The local search is based on simulated annealing with the logarithmic cooling schedule $c(k) = \Gamma / \ln(k+2)$. The parameter Γ depends on S and the neighbourhood relation is determined by the classical Perceptron algorithm. The examples are fragments of DICOM CT images of size $n = 14161 = 119 \times 119$. On test sets of $50 + 50$ examples (disjoint from the learning set) we obtain a correct classification of about 97%.

1 Introduction

Advances in modern imaging technology, in particular in sectional imaging modalities (computed tomography and magnetic resonance imaging), produce an ever increasing amount of data to be assessed by the radiologist. New techniques of data postprocessing are therefore used to assist in diagnosis and presentation of the findings. Automated preselection of the up to 1500 slices acquired in a single examination and their classification as normal or abnormal findings considerably facilitates image interpretation and reduces physician time. The present study was performed to investigate the computer-assisted interpretation of CT scans of the liver by depth-three threshold circuits.

Forty cases with normal findings and 150 cases with liver pathology were selected for the study from a total of 735 abdominal and biphasic upper abdominal CT examinations performed with comparable imaging parameters. The abnormal findings were confirmed by histology in 83 patients and by follow-up in patients with known malignancy in 67 instances. Ten to fifteen slices were selected for the cases with normal findings and between one and five slices from the portal-venous phase for those with abnormal findings. Representative areas from these slices depicting normal liver tissue as well as vessel cross-sections and cysts or one or more focal liver lesions surrounded by normal liver tissue were identified and recalculated for a matrix of 119×119. A total number of 348 ROIs showing normal liver parenchyma and another 348 ROIs with abnormal findings (hypodense lesions) were used to calculate the depth-three circuits.

We utilise a new method to compute depth-three threshold circuits from sample sets by a combination of the classical Perceptron algorithm with simulated annealing. For sample sets S of n-dimensional vectors $\mathbf{x}$ that are separable by a linear threshold function into "positive" and "negative" examples, MINSKY and PAPERT [9] proved the following convergence property: If $\mathbf{w}^*$ is a unit vector solution to the separation problem, then the Perceptron algorithm converges in at most $1/\sigma^2$ iterations, where $\sigma := \min_{[\mathbf{x},\eta]\in S} |\mathbf{w}^* \cdot \mathbf{x}|$, $\eta \in \{+,-\}$. The parameter σ can be exponentially small in terms of the dimension n.

In general, the simple Perceptron algorithm performs well even if the sample set is not consistent with any weight vector $\mathbf{w}$ of linear threshold functions, see [6]. For our problem of CT image classification, one can hardly assume that positive and negative examples are separable by a single linear threshold function. In order to reduce the classification error, we try to compute a bounded-depth circuit consisting of linear threshold functions. The threshold functions, in particular the gates of the first level, are determined by a learning procedure from positive and negative examples S of the classification problem.

HÖFFGEN [8] has shown that finding a linear threshold function that minimises the number of misclassified examples is NP-hard in the case of arbitrary sample sets. We approach this computationally hard minimisation problem by a simulated annealing-based procedure where the neighbourhood relation is determined by the Perceptron algorithm. The combination has been studied in [3] for samples generated by non-linear threshold functions.

To our knowledge, the first paper on learning-based methods applied to X-ray diagnosis was published by ASADA ET AL. [4]. Since then, the research has been concentrating on using commercially available neural networks for medical image classification [5, 7, 10, 11].

In a number of papers, feature extraction is used in learning-based classification methods [10, 12]. In [10], for example, a high classification rate of nearly 98% is reported, where the Wisconsin breast cancer diagnosis (WBCD) database of 683 cases is taken for learning and testing. The approach is based on feature extraction from image data and uses nine visually assessed characteristics for learning and testing. Among the characteristics are the uniformity of cell size, the uniformity of cell shape, and the clump thickness.

In our approach, the input to the algorithm are fragments of CT images of size 119×119 with an 8 bit grey scale in DICOM standard format. Therefore, the input size is $n = 14161$ and the input values range from 0 to 255. For the learning procedure, we used 348 positive (abnormal findings) and 348 negative (normal liver parenchyma) examples. The result of the algorithm is a depth-three threshold circuit consisting of 37 gates (linear threshold functions). The circuits have been tested on $50 + 50$ examples (different from the learning set), and we obtained a correct classification of about 97%. The time to compute the depth-three circuit is about 30 hours, the classification itself (i.e., the test) is performed within a few seconds.

2 Simulated Annealing and the Perceptron Algorithm

We assume that rational numbers are represented by pairs of binary tuples of length d and denote the set of linear threshold functions by

$$\mathcal{F} := \bigcup_{n\geq 1} \mathcal{F}_n, \text{ where } \mathcal{F}_n = \{\, f(\mathbf{x}) \,:\, f(\mathbf{x}) = \sum_{i=1}^{n} w_i \cdot x_i \geq \vartheta_f \,\},$$

where w_i and x_i are equal to $\pm(p_i, q_i)$ for $p_i,\, q_i \in \{0,1\}^d$.

The functions from $\mathcal{F}$ are used to design single-output circuits $\mathcal{C}$ of threshold functions: A circuit $\mathcal{C}$ is defined by the underlying acyclic directed graph $\mathcal{G} = [E, V]$, $E \subset V \times V$. The graph $\mathcal{G}$ has n input nodes labelled by variables x_1, $\cdots$, x_n, and $\mid V \mid - n$ nodes v_f labelled by threshold functions $f \in \mathcal{F}$, where the number of incoming edges of v_f has to be consistent with the number of variables of f. Finally, one v_f is chosen as the output v_{out} of $\mathcal{C}$.

The depth of $\mathcal{C}$ is the maximum number of edges on a path from an input node x_i to the output node v_{out}. The nodes that are not input nodes are called gates. The function $F(\mathcal{C})$ computed by $\mathcal{C}$ is defined as follows: The gates of the first level output 1 or 0 depending on whether or not $\sum_{i=1}^{n} w_i \cdot x_i \geq \vartheta_f$. In the same way, the gates at higher levels have Boolean outputs only. Therefore, when all paths from input nodes to v_{out} are of the same length, the gates at level 2, 3, .. do compute Boolean threshold functions. Thus, we have $F(\mathcal{C}) : \{0,1\}^{n\cdot d} \rightarrow \{0,1\}$.

In the present paper, the maximum depth of $\mathcal{C}$ is three; circuits of depth one are simply the elements of $\mathcal{F}_n$, and in Section 3 we consider circuits of depth two and three, respectively. In our application, each of the threshold functions from the first level is equally important for the overall classification result of the depth-three circuit. Therefore, the weights at the input lines of second level functions (gates) can be normalised to the value 1 and only the threshold values depend on the sample set S.

For a given sample set S, we assume $S = \{\,[\mathbf{x}, \eta]\,\}$ for $\eta \in \{+, -\}$ and $\mathbf{x} = (x_1, ..., x_n)$ where $x_i = (p_i, q_i)$, $p_i,\, q_i \in \{0,1\}^d$. Furthermore, we consider a particular number n of variables only and we take the set $\mathcal{F} := \mathcal{F}_n$ as the configuration space.

The objective of our optimisation procedure is to minimise the number $\mid S\Delta f \mid$ of misclassified examples, $S\Delta f := \{[\mathbf{x}, \eta] : f(\mathbf{x}) < \vartheta_f \& \eta = + \text{ or } f(\mathbf{x}) > \vartheta_f \& \eta = -\}$, and we denote $\mathcal{Z}(f) := \mid S\Delta f \mid$.

Given $f = \sum_{i=1}^{n} w_i \cdot x_i \geq \vartheta_f$, the neighbourhood relation $\mathcal{N}_f$ is suggested by the Perceptron algorithm and defined by

$$w_i(f') := w_i \,-\, y_j \cdot x_{ij} / \sqrt{\sum_{i=1}^{n} w_i^2},\; j \in \{1,\, 2,\, \ldots, m\}, \tag{1}$$

for all i simultaneously and for a specified j that maximises $\mid y_j - \vartheta_f \mid$, where $y_j = \sum_{i=1}^{n} w_i \cdot x_{ij}$. The threshold $\vartheta_{f'}$ is equal to $\vartheta_f + y_j / \sqrt{\sum_{i=1}^{n} w_i^2}$.

Given a pair $[f, f']$, $f' \in \mathcal{N}_f$, we denote by $G[f,\, f']$ the probability of generating f' from f and by $A[f, f']$ the probability of accepting f' once it has been generated from f. To speed up the local search for minimum error solutions, we take a non-uniform generation probability where the transitions are forced into the direction of the maximum deviation (we used a similar approach in [1]).

The non-uniform generation probability is derived from the Perceptron algorithm: When f is the current hypothesis, we set

$$U(\mathbf{x}) := \begin{cases} -f(\mathbf{x}), & \text{if } f(\mathbf{x}) < \vartheta_f \text{ and } \eta(\mathbf{x}) = +, \\ f(\mathbf{x}), & \text{if } f(\mathbf{x}) \geq \vartheta_f \text{ and } \eta(\mathbf{x}) = -, \\ 0, & \text{otherwise.} \end{cases} \qquad (2)$$

For $f' \in \mathcal{N}_f$, we set $G[f,\, f'] := U(\mathbf{x}) / \sum_{\mathbf{x} \in S\Delta f} U(\mathbf{x})$. Thus, preference is given to the neighbours that maximise the deviation. Now, our heuristic can be summarised in the following way:

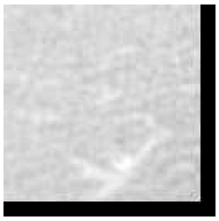
1. The initial hypothesis is defined by $w_i = 1$, $i = 1, 2, \ldots n$ and $\vartheta = 0$.
2. For the current hypothesis, the probabilities $U(\mathbf{x})$ are calculated; see (2).
3. To determine the next hypothesis f_k, a random choice is made among the elements of $\mathcal{N}_{f_{k-1}}$ according to the definition of $G[f,\, f']$.
4. When $\mathcal{Z}(f_k) \leq \mathcal{Z}(f_{k-1})$, we set $A[f_{k-1}, f_k] := 1$.
5. When $\mathcal{Z}(f_k) > \mathcal{Z}(f_{k-1})$, a random number $\rho \in [0, 1]$ is drawn uniformly.
6. If $A[f_{k-1}, f_k] := e^{-(\mathcal{Z}(f_k) - \mathcal{Z}(f_{k-1})/c(k)} \geq \rho$, the function f_k is the new hypothesis. Otherwise, we return to 3 with f_{k-1}.
7. The computation is terminated after a predefined number of steps K.

Hence, instead of following unrestricted increases of the objective function, our heuristic tries to find another "initial" hypothesis when the difference of the number of misclassified examples is too large.

The crucial parameter $c(k)$ is defined by $c(k) = \Gamma / \ln(k+2)$, $k = 0, 1, \ldots$ When Γ is larger than or equal to the maximum value of the minimum escape depth from local minima, one can prove the convergence to minimum-error solutions for a more general neighbourhood relation that provides the reversibility of $\mathcal{F}$. In this case, the convergence analysis from [2] indicates a time complexity of roughly $n^{\Gamma + O(1)}$, i.e., after $n^{\Gamma} + \log^{O(1)}(1/\delta)$ transitions the confidence that a minimum-error threshold function has been computed is larger than $1 - \delta$.

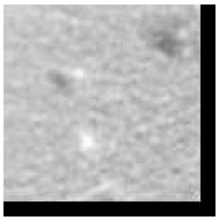
3 Computational Experiments

The heuristic was implemented in C^{++} and we performed computational experiments on SUN Ultra 5/333 workstations with 128 MB RAM.



Normal liver tissue: Negative example.

Figure 1



Focal liver tumour: Positive example.

Figure 2

In the experiments, we used fragments of CT images of size 119×119 with 8 bit grey levels. From 348 positive (with abnormal findings) and 348 negative examples (normal tissue) several independent hypotheses of the type $w_1 \cdot x_1 + \cdots + w_n \cdot x_n \geq \vartheta$ were calculated for $n = 14161$. We tested the hypotheses on 50 positive and 50 negative examples. The test examples were completely different from the learning set.

Table 1 summarises typical results for circuits of depth 1, ..., 3. Each function (gate) from the first level was trained on a random choice of $50+50$ examples out of $348+348$ examples. The examples were learned with zero error when $\Gamma \geq 10$ (Table 1 is for $\Gamma = 15$). The depth-three circuit consists of three sub-circuits of depth two, where each depth-two circuit has 11 threshold functions at the first layer. The output gate of the depth-three circuit is a simple majority function. Thus, the depth-three circuit consists of $3 \cdot (11+1)+1 = 37$ gates.

Depth of Circuits	Learning Run-Time	Errors on POS	NEG	Errors on T_POS	T_NEG	Percentage of Errors
1	49 min	0	0	13	16	29%
2	537 min	0	0	3	5	8%
3	1669 min	0	0	1	2	3%

Table 1

Each of the threshold functions of the first level (i.e., each input gate) has $n = 14161$ inputs, i.e., the total number of input lines that are connected to the 14161 input nodes (pixel values) is $3 \cdot 11 \cdot 14161 = 467313$.

References

1. A. Albrecht, S.K. Cheung, K.S. Leung, and C.K. Wong. Stochastic Simulations of Two-Dimensional Composite Packings. *J. of Comput. Physics*, 136:559–579, 1997.
2. A. Albrecht and C.K. Wong. On Logarithmic Simulated Annealing. In: J. van Leeuwen, O. Watanabe, M. Hagiya, P.D. Mosses, T. Ito, eds., *Theoretical Computer Science: Exploring New Frontiers of Theoretical Informatics*, pp. 301 – 314, LNCS Series, vol. 1872, 2000.
3. A. Albrecht and C.K. Wong. Combining the Perceptron Algorithm with Logarithmic Simulated Annealing. To appear in: *Neural Processing Letters*.
4. N. Asada, K. Doi, H. McMahon, S. Montner, M.L. Giger, C. Abe, Y.C. Wu. Neural Network Approach for Differential Diagnosis of Interstitial Lung Diseases: A Pilot Study. *Radiology*, 177:857 – 860, 1990.
5. D.B. Fogel, E.C. Wasson III, E.M. Boughton and V.W. Porto. Evolving Artificial Neural Networks for Screening Features from Mammograms. *Artificial Intelligence in Medicine*, 14(3):317, 1998.
6. S.I. Gallant. Perceptron-Based Learning Algorithms. *IEEE Trans. on Neural Networks*, 1(2):179 – 191, 1990.
7. H. Handels, Th. Roß, J. Kreusch, H.H. Wolff and S.J. Pöppl. Feature Selection for Optimized Skin Tumour Recognition Using Genetic Algorithms. *Artificial Intelligence in Medicine*, 16(3):283 – 297, 1999.
8. K.-U. Höffgen. Computational Limitations on Training Sigmoid Neural Networks. *Information Processing Letters*, 46(6):269 – 274, 1993.
9. M.L. Minsky and S.A. Papert. *Perceptrons.* MIT Press, Cambridge, Mass., 1969.
10. C.A. Pea-Reyes and M. Sipper. A Fuzzy-genetic Approach to Breast Cancer Diagnosis. *Artificial Intelligence in Medicine*, 17(2):131 – 155, 1999.
11. A.L. Ronco. Use of Artificial Neural Networks in Modeling Associations of Discriminant Factors: Towards an Intelligent Selective Breast Cancer Screening. *Artificial Intelligence in Medicine*, 16(3):299 – 309, 1999.
12. C. Roßmanith, H. Handels, S.J. Pöppel, E. Rinast, and H.D. Weiss. Computer-Assisted Diagnosis of Brain Tumors Using Fractals, Texture and Morphological Image Analysis. In: H.-U. Lemke, ed., *Proc. CAR*, pp. 375–380, 1995.