On the use of Neural Networks in Statistical Shape Analysis

Stefan Giebel[1], Jens-Peter Schenk[2], Jang Schiltz[1]

[1] Luxembourg School of Finance
University of Luxembourg, Luxembourg
(e-mail: Jang.Schiltz@uni.lu)
[2] Pädiatrische Radiologie
Universittsklinik Heidelberg, Heidelberg, Deutschland
(e-mail: Jens-peter.Schenk@med.uni-heidelberg.de)

Abstract. There are different kinds of tumours that can appear in childhood: nephroblastoma, clear cell sarcoma, neuroblastoma etc. The chosen therapy depends upon the diagnosis of the radiologist which is done with the help of MRI (Magnetic resonance images). Our research is the first mathematical approach on MRI of renal tumours (n=80). We are using transversal, frontal and sagittal images and compare their potential for differentiation of the different kind of tumours by use of Statistical Shape Analysis. We determine the key points or three dimensional landmarks of the renal tumours by using the edges of the platonic body (C60). Furthermore we use a combination of Neural Networks and Statistical Shape Analysis to consider all kinds of linear transformations and compare the results to the one obtained by the traditional test of Ziezold test for the determination and differentiation of the mean shape.
Keywords: Neural Networks, Statistical Shape Analysis, Mean Shape, Renal tumours.

1 Introduction

In a wide variety of disciplines it is of great practical importance to measure, describe and compare the shapes of objects. In general terms, the shape of an object, data set, or image can be defined as the total of all information that is invariant under translations, rotation and isotropic rescaling. The field of shape analysis involves hence methods for the study of the shape of objects where location, rotation and scale can be removed. The two- or more dimensional objects are summarised according to key points called landmarks. This approach provides an objective methodology for classification whereas even today in many applications the decision for classifying according to the appearance seems at most intuitive.
Interest in shape analysis began in 1977. D.G. Kendall[7] published a note in which he introduced a new representation of shapes as elements of complex projective spaces. K.V. Mardia[10] on the other hand investigated the distribution of the shapes of triangles generated by certain point processes, and in particular considered whether towns in a plain are spread regularly with

equal distances between neighbouring towns. The full details of this elegant theory which contains interesting areas of research for both probabilists and statisticians where published by D. Kendall[8] and F. Bookstein[1]. The details of the theory and further developments can be found in the textbooks by C.G. Small[14] and I.L. Dryden & K.V. Mardia[4].
The renal tumour is limited by spleen or liver , the rest of the kidney, the spine and retroperitoneal vessels. In Giebel(2007)[2] it was shown that every landmark has another meaning for differentiating the tumours. Giebel et al. [3] showed that none of the landmarks has a special influence for the determination of the mean shape according to the test of Ziezold (2003)[17].
In this paper, the edges of the platonic body (C60) define the landmarks. We use a combination of Neural Networks and Statistical Shape Analysis and compare the results to the one obtained by the traditional test of Ziezold test for the determination and differentiation of the mean shape.

2 Wilms' tumours

Nephroblastoma (Wilms' tumour)[15] is the typical tumour of the kidneys appearing in childhood. Therapy is organised in therapy-optimizing studies of the Society of Paediatric Oncology and Haematology (SIOP). Indication of preoperative chemotherapy is based on radiological findings. The preferred radiological method are sonography and MRI. Both methods avoid radiation exposure, which is of great importance in childhood. Preoperative chemotherapy is performed without prior biopsy[12].
Information of the images of magnetic resonance tomography, especially the renal origin of a tumour and the mass effect with displacement of other organs, is needed for diagnosis. Next to nephroblastomas other tumors of the retro peritoneum exist, which are difficult to differentiate [13]. Renal tumours in childhood are classified into three stages of malignancy (I, II, III). Typical Wilms tumors mostly belong in stage II. In stage II different subtypes of nephroblastoma tissue exist[6].
In our sample of tumours, four different types of retroperitoneal tumours are represented: nephroblastoma, neuroblastoma, clear cell carcinoma, and renal cell carcinoma. Renal cell carcinomas are very rare in childhood. They represent the typical tumours of adult patients. They have no high sensitivity for chemotherapy. Clear cell sarcomas are very rare in childhood and are characterised by high malignancy. Neuroblastomas are the typical tumours of the sympathetic nervous system and suprarenal glands. Infiltration of the kidney is possible.
The tumour grows with encasement of vessels. Because of the high importance of radiological diagnosis for therapy, it is of great interest to find markers for a good differentiation of tumours. MRI produces 2D-images. From the two dimensional data a three dimensional object has to be computed.Image 1 shows an example of the raw data.

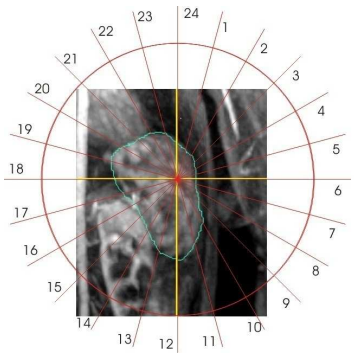

Fig. 1. 2D-image of the tumour

3 Mean Shape

To compare the standardised and centred sets of landmarks, we have to define the mean shape of all the objects and a distance function which allows us to evaluate how "near" every object is from this mean shape.
The term "mean" is here used in the sense of Fréchet (1948)[5]. If X denotes a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ with values in a metric space (Ξ, d), an element $m \in \Xi$ is called a mean of $x_1, x_2, ..., x_k \in \Xi$ if

$$\sum_{j=1}^{k} d(x_j, m)^2 = \inf_{\alpha \in \Xi} \sum_{j=1}^{k} d(x_j, \alpha)^2. \tag{1}$$

That means that the "mean shape" is defined as the shape that garantees the smallest possible variance for a group of objects. For computing the mean shape we use the algorithm of Ziezold (1994)[16].
In the special case of oncology there is no theoretical medical reason to select a specific group of landmarks for differentiation. All landmarks in this research have thus to be selected by an explorative procedure.
The test of Ziezold (1994)[16] is a statistical test which allows to determine if a given object belongs to a set of objects defined by their mean shape. We have used this test to see if given Wilm's tumoursw can be differentiated from the mean shape of the neuroblastomas and vice versa.

4 Elements of neural networks

Neural networks have been developed originally in order to understand the cognitive processes. Nowadays there are a lot of applications of neural networks as a mathematical method in various quite different disciplines. The term "neural networks" points to the model of a nerve cell, the neuron,

and the cognitive processes carried and driven by the network of interacting neurons. A neuron perceives chemical and physical excitement from the environment by its dendrites. The neuron is processing this incoming data and sending the information to other neurons via axons and synapses. McCulloch and Pitts implemented the biological processes of a nerve cell for the first time in a mathematical way [9]. Nerve cells have to access and process incoming data in order to evaluate target information. Therefore the corresponding neural networks are called supervised neural networks.
An unsupervised neural network has no target and is similar to a cluster algorithm. The data consist of n variables $x_1, \ldots, x_n$ on binary scale. For data processing, the ith variable x_i is weighted with w_i. Normalised with $|w_i| \leq 1$, multiplication of x_i with w_i determines the relevance of x_i for a target y. The value w_i reflects the correlation between the input variable and the target, the sign indicating the direction of the influence of the input variable on the target. Weighting the input variables for a target variable is similar to discriminant analysis.The critical quantity for the neuron is the weighted sum of input variables

$$q := \sum_{i=1}^{n} w_i \cdot x_i = w_1 \cdot x_1 + ... + w_n \cdot x_n \quad . \tag{2}$$

For a target y with binary scale, a threshold S is needed. Crossing the threshold yields 1 and falling below the threshold yields 0. Hence the activation function F can be written as

$$F(q) = \begin{cases} 1, & \text{if } \; x > S \\ 0, & \text{if } \; x \leq S \end{cases} \tag{3}$$

In comparison to discriminant analysis, for neural networks the threshold S has to be assigned, depending on properties of the target; it can not be derived from the data in a straightforward manner. Neural networks usually include no assumption about the data, they are a purely numerical method.

With the input (2) of the activation function, we obtain $y = F(q)$ as

$$y = 1, \quad \text{if } \; \sum_{i=1}^{n} w_i \cdot x_i > S$$

$$y = 0, \quad \text{if } \; \sum_{i=1}^{n} w_i \cdot x_i \leq S$$

Multi-layer neural networks are able to solve all logical functions for separating groups.

5 Multi-layer perceptrons

In general a given target may be reached only up to a certain error. Given a certain measure $E(\widetilde{y}, y)$ for the distance between the given target state y

and the state $\tilde{y}$ computed by the neural network, the learning of the neural network corresponds to the minimisation of $E(\tilde{y}, y)$. The following training algorithm is inspired by Rumelhart, Hinton and Williams [11]. The total error measure over all states of a given layer is defined as

$$E_{total}(\tilde{y}, y) := \frac{1}{2}\sum_{k=1}^{N}(\tilde{y}_k - y_k)^2 \quad . \tag{4}$$

It will be used below to reset the weights in each layer of the neural network.

The processed state $\tilde{y}$ of the neural network is computed by the following steps.

First the critical parameter for the first layer is computed from n weighted input values as $\sum_{i=1}^{n} w_i \cdot x_i$. We consider a hidden output layer with m neurons. For $j = 1, \ldots, m$, let g_j be the activation function of the j-th neuron of the hidden layer, with an activation value of h_j, given as

$$h_j = g_j(\sum_{i=1}^{n} w_i \cdot x_i) \quad . \tag{5}$$

Usually for all neurons of a given layer a common activation function $g = g_1, \ldots, g_m$, e.g. a sigmoid function, is used.

Next, the output of the previous (hidden) layer becomes the input of the next layer, and the activation proceeds analogously to the previous layer. Let f be the activation function of the pre-final (here the second) output layer. Then the pre-final critical value is

$$q = f(\sum_{j=1}^{m} u_j \cdot h_j) \quad . \tag{6}$$

Finally, the pre-final critical value q is interpreted by a final activation function F yielding

$$\tilde{y} = F(q) \tag{7}$$

as a final state value computed from the neural network with the given weights of the input variables from input and hidden layers. Now the neural network performs a training step by modifying the weights of all input layers. The learning mechanism the weights is determined by the target distance measure

$$E = \frac{1}{2}\sum_{i=1}^{n}(y^i - \tilde{y}^i)^2 \quad .$$

The weights of both layers are changed according to the steepest descent, i.e.

$$\Delta w_i = \frac{\partial E}{\partial w_i} \tag{8}$$

$$\Delta u_j = \frac{\partial E}{\partial u_j} \tag{9}$$

With a learning rate α, which should be adapted to the data, the weights are changed as follows:

$$w_i^{new} = w_i^{old} - \alpha \cdot \Delta w_i \tag{10}$$

$$u_j^{new} = u_j^{old} - \alpha \cdot \Delta u_j \tag{11}$$

The necessary number of iterations depends on the requirements imposed by the data, the user, and the discipline.
For simplicity, we consider now an 1-layer perceptron network, which is sufficient for our purpose of minimising the variance. Every landmark is weighted in every direction.

$$\sum_{j=1}^{k} d(x_j, m)^2 = \inf_{\alpha \in \Xi} \sum_{j=1}^{k} d(x_j, \alpha)^2. \tag{12}$$

In contrast to the former application of neural networks we are using a metric function instead of a binary variable. The difference between the weighted objects and the approximated mean shape is used instead of the difference between the reality and the approximation E.

6 Results

To get $3D$ landmarks we construct a three dimensional object of the tumour from the $2D$ MRI. Then we take the intersection between the surface of the tumour and the vectors going from the centre to the edges of the platonic body C60 as landmarks as is shown in figure 2.

Minimising the variance in one of the groups does not lead always to an

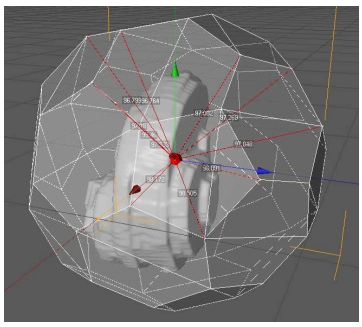

Fig. 2. 3D-Landmarks as cut points between the edge of a platonic body and the surface of the tumor

optimal differentiation between the different types of tumors. The neuronal network uses in fact a different metric to minimising the variance. Every

landmark is weighted in every direction. For a sample of 74 comparable tumors (69 nephroblastoma and 5 neuroblastoma the u_0-values are computed for comparing our nephroblastomas to the mean shape of neuroblastomas.

The range of the u_0-values computed by the MLP lies between 0 and 188. If use the the test of Ziezold[16] with the Euclidean distance instead of the distance applied in MLP, we get an u_0-value of 112. For a randomised sample ($n = 1000$), we get a p-value of 0.080.
If we compare our neuroblastomas to the mean shape of nephroblastomas, we get an u_0 value of 72 with a p-value of 0.116 in a randomised sample ($n = 1000$).
Figure 3 shows the mean shape of the nephroblastomas (red) and of the neuroblastomas (green).

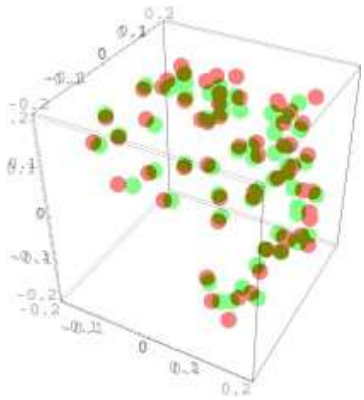

Fig. 3. Mean Shape: Red: 60 landmarks of the mean shape of the nephroblastoma, Green: 60 landmarks of the mean shape of the neuroblastoma

7 Conclusion

The neuroblastoma can be differentiated quite well from the mean shape of the nephroblastoma, especially if we use the Euclidian distance as metric. Shape Analysis is useful to make a decision in spite of different size, location etc. The test used for differentiating the existing kind of tumours does not need any assumptions in regard to distributions and the size of the sample.
For improving our results we will try to use approriate non-Euclidean transformations in the neural networks. A possible approach is to use a supervised 2-layer neural network with weighted landmarks. We will minimise the variance to estimate a "mean shape" in one of the groups instead of minimising

the mistake between output and reality. Indeed, we have seen that a small variance does not always allow an optimal differentiation between the groups. Not every transformation leads to a better differentiation of tumours. If the size or location of tumours plays a role in differentiation, it could be wrong to centre or standardise the objects.

References

1. Bookstein, F.L., "Size and shape spaces for landmark data in two dimensions", *Statistic Sciences* 1, 181-242 (1986).
2. Giebel, S., *Statistical Analysis of the shape of renal tumors in childhood*, Diploma thesis, University of Kassel (2007).
3. Giebel, S., Schiltz, J., and Schenk, J.P., "Application of Statistical Shape Analysis on renal tumours in 3D" *Bulletin de la Soicété des Sciences Médicales du Grand-Duché du Luxembourg* to appear (2010).
4. Dryden, I.L., and Mardia, K.V., *Statistical Shape Analysis* Wiley, Chichester (1998).
5. Fréchet, M., "Les élements aléatoires de nature quelconque dans un espace distancié", *Annales de l'Institut Henri Poincaré*, 10, 215-310 (1948).
6. Graf, N., *Urologe* A 43:421 (2003).
7. Kendall, D.G., "The diffusion of shape", *Adv. Appl. Probab.* 9 (1977)
8. Kendall, D.G., "Shape manifolds, Procrustrean metrics and complex projective spaces", *Bulletin of the London Mathematical Society*, 16, 81-121 (1984).
9. McCulloch, W.S., and Pitts, W., "A logical calculus of the idea immanent in neural nets", *Bulletin of Mathematical Biophysics*, 115-137 (1943).
10. Mardia, K.V., "Mahalanobis distance and angles", In: Krishnaiah, P.R. (ed) *Multivariate Analysis IV*, 495-511, North Holland, Amsterdam (1977).
11. Rumelhart, D.E., Hinton, G.E., and Williams, R.J., *Learning internal representation by error propagation*, 318-362 (1986).
12. Schenk, J.P, *Fortschr. Röntgenstr.* 178:38 (2006).
13. Schenk, J.P, *Eur. Radiol.* 18:683 (2008).
14. Small, C. G., *The Statistical Theory of Shape*, Springer-Verlag, New York (1996).
15. Wilms, M., *Die Mischgeschwülste der Niere*, Verlag von Arthur Georgi, Leipzig, 1-90 (1889).
16. Ziezold, H., "Mean Figures and Mean Shapes Applied to Biological Figure and Shape Distributions in the Plane", *Biometrical Journal* 36, 491-510 (1994).
17. Ziezold, H., "Independence of Landmarks of Shapes", *Mathematische Schriften Kassel*, Heft 03 (2003).