Analysing Hierarchical Data Using a Stochastic Evolutionary Neural Tree

R.G.Adams, N.Davey, S.J.George

R.G.Adams@herts.ac.uk, N.Davey@herts.ac.uk, S.J.George@herts.ac.uk
Faculty of Engineering and Information Sciences,
University of Hertfordshire,
Hatfield, Herts., UK. AL10 9AB
Tel +44 01707 284321

Abstract

SCENT is simple competitive neural network model that evolves a tree structured set of nodes in response to being presented with an unlabelled data set. The resulting set of weight vectors and their relationship can be viewed as giving a hierarchical classification of the training data. This paper examines the nature of this classification for two data sets over several runs of the network. The first data set is a set of grey scale images, chosen because the code-vectors produced by SCENT can then be visualised in a natural way. The second data set is a small set of vectors coding attributes of animals. The resulting taxonomy from SCENT can then be compared with the normal taxonomic groups that such a set of animals would fall into. Since the SCENT model is stochastic different runs produce different trees, but the variation in results produced over several runs is small. The model is shown to be reasonably robust and the relationship between the nature of the data and the type of tree produced is examined.

1 Introduction

This paper describes how hierarchical structure can be discovered in unlabelled data using a stochastic evolutionary neural tree model (SCENT), developed at the University of Hertfordshire. The data that we present to the net in these experiments has no explicit hierarchical structure, but both data sets have been chosen to allow various aspects of the resulting hierarchies to be examined.

Neural networks may be structured to reflect aspects of their training data. The benefits of imposing a topology onto the nodes of a competitive network are well documented. For example a self organising map (SOM) may organise the nodes into a two dimensional grid, and other models, including SCENT, organise the nodes into a tree structure [Butchart et al., 1995]. SCENT is an evolutionary model, in which nodes are added and deleted in response to the data being classified; the tree structure is a natural way of representing the way in which nodes grow, producing children. Moreover the resulting tree structure provides hierarchical information about the data, as well as the conventional codevector clustering present in the leaf nodes of the tree. The vectors clustered by a node in the tree can be interpreted as being recursively subclustered by the children of that node. Just as the weight vectors of leaf nodes provide bottom level code-vectors the higher level nodes give representations of the centroids of their corresponding clusters.

The model is briefly described in section 2, and the performance of SCENT on two data sets is then reported in the next two sections. The final analysis section summarises the overall performance of SCENT and concludes.

2 SCENT

Detailed descriptions are given [Butchart et al., 1997] so here only a summary of the model is given.

Learning in SCENT takes place by a simple recursive search through the tree, with the winner in each subtree, the node whose weight vector is nearest in Euclidean space, moving towards the current input vector. As in other neural clusterers the movement is mediated by a learning rate.

At initialisation SCENT consists of a single root node that will eventually move to the centroid of the data. As data vectors are presented to the net, nodes may produce growth. A node is allowed to grow when its relative activity, the ratio of the number of times it has won to the number of times its parent has won, exceeds a threshold. If a node has classified vectors that are relatively near it, so that it is representing a dense cluster, then it produces two children to subclassify the cluster. On the other hand if the data it is classifying is spatially separated then the growth produces a sibling. In either case any new node is a noisy copy of the original.

To fully explore the data it is desirable for the network to create much tentative growth which may later be pruned. If a node does not have a classificatory error significantly less than its parent, or is insufficiently active relative to its depth in the tree then it is removed. Pruning is stochastic in SCENT. In early epochs pruning is more likely than later in the classificatory process when less growth is taking place. A form of simulated annealing controls the probability of pruning. Typically only 20% of all the nodes created will still be in existence in the final tree.

SCENT is a fully autonomous model. The net is given no indication of the appropriate number of classifying nodes to use, nor any indication of a desirable hierarchical structure. The results presented below use identical runtime initialisations of the program for both data sets. However each run of SCENT produces a unique tree, so that we present results over several runs to address issues of stability and repeatability.

As a simple codevector classifier of unlabelled data the performance of SCENT is comparable to the better unsupervised neural net clusterers, such as Neural Gas [Martinetz et al., 1993], see [Butchart et al., 1996] for a full comparison. Here, though, we investigate the nature of the hierarchical structure produced for two data sets where the vectors have a clear semantic and a meaningful structure.

3 The Picture Data

The first data set consists of 153 vectors [Gale 1997]. Each vector has 2500 grey scale elements in the range 0-255 representing a 50*50 grid of pixels. The vectors have been formed by scanning a set of pictures with care being taken to ensure evenly sized and centralised images. There is no semantic information included in the input vectors, which consist purely of the grey scale images. There are 17 categories of image, such as snakes, birds, fish, clocks, tables and guitars. There are 3 varieties of each category with the exception of pianos (with 2) and cabbages/lettuces (with 4) giving 51 different images, each image occurs in 3 different contrast variants. The high contrast set is shown in Figure 1.

In the grey scale images white is represented as 255 and black is 0. As can be seen from the set of images some are lighter in general tone (fish, birds, mice etc..) and some are generally darker (armchair, cabbages, upright piano etc.). The lighter images are in general smaller varying from about 10% of the image (250 pixels) to about 40% of the image (1000 pixels). The darker images are generally larger varying from about 20% to 80% of the image.

3.1 Results

Figure 2 shows a typical result formed by using SCENT on the 153 vector data set. Each final and non final node is represented in the tree by its classifying weight vector converted back into a 50*50 pixel image. It generally collects the 3 contrast variants of the same image together despite the fact that they are presented in a random order each epoch. The reason for this can be seen by considering the metric used by SCENT to determine the nearest vector for classification purposes. This metric is the Euclidean distance formula:

$$d = \sqrt{\sum_{i=0}^{2499} (x_i - w_i)^2}$$

where d is distance, x_i is the ith component of the input vector and w_i is the ith component of the weight vector.

In the lighter images the darkest contrast variants average pixel values around 125 and the lightest contrast variants average around 175. In the darker images the darkest contrast variants average pixel values around 50 and the lightest contrast variants average around 90.

From this it can be seen that, since white is represented as 255, the difference in vector component for an image with a non-white pixel as against a white pixel is larger than the difference in vector component for two contrast variants of the same image, considerably larger in the case of a dark contrast darker image. Hence two different images will generally have a larger distance from each other than will two different contrast variants of the same image.

These pixel values also explain the ability of the model to separate lighter and darker images, since for each of the dark images the difference between a white pixel with no image and a pixel with an image is considerable (up to 80% of the component's possible contribution) whereas a lighter image is generally smaller and each pixel is less different from white.

Figure 2 also illustrates one of the benefits of using this set of data. The visual character of the data makes it possible to see the nature of the analysis provided by the network. It can be seen that the weight vectors move from rough generalisations at the top of the tree structure to more detailed pictures at the lower levels. In general the tree splits the lighter and darker images into separate sub-trees. The left hand sub-tree has classified the lighter and smaller images and the other two sub-trees classifying darker and larger images with the right hand sub-tree classifying the darkest images.

Within the left hand sub-tree the next level classifies according to the general orientation of the image. The first and second classifying vertical images; the first

classifies all the spiders, the vertical snake (adder), a floor lamp and a more vertically oriented mouse; the second deals with the darker variants - all the guitars, the grandfather clock, the office chair and a rogue potato. The third classifies the more diagonal elements, both deer and a snake. Finally the fourth classifies the horizontal images, all the birds, all the fish, a snake and two of the mice.

In the middle sub-tree the first and second sub-part are leaf nodes. The first classifies slightly more horizontal elements, a coffee table, digital clock and a bed, while the second classifies ones with more dark areas around the top half, a drum and two tables. The final one has more general mid-dark images including virtually all the frogs and the elk, which is darker than the two deer and so is in this middle sub-tree.

In the right hand sub-tree we get all the really dark images. All the cabbages are in this sub-tree together with large beds and pianos and the dark chairs and potatoes. Of the three sub-classifications the middle one has the more vertical images the right hand one is a leaf node with the dark areas slightly to the top of the image and the left hand one has the largest dark images.

In two cases two leaf nodes appear to be the same as their parent node. This is partly because the parent nodes are classifying two images and the full sized images show more of a difference and also that the child nodes are obviously newly created and have not yet had time to differentiate properly.

SCENT has succeeded in producing a hierarchical tree structure reflecting the clear visual differences between the images.

4 The Zoo Data

The Zoo machine learning database [Murphy & Aha, 1992] is used as the second data set. It contains 101 vectors which describe 101 instances of animals in terms of 18 attributes: a naming label “animal name”; 15 binary attributes such as: aquatic or venomous; a field giving the number of legs, an integer between 0 and 8; and a type label. The type label indicates a biological taxonomic group dividing the 101 instances into 7 categories of animal, each uniquely labelled by an integer. Six of the categories clearly define large taxonomic groups, such as: mammals, birds and insects. The seventh categorises other animals: clams, crabs, crayfish, lobsters, starfish, octopus, scorpion, seawasp, slug and worm. An example input vector for the platypus is presented in Table 1, below.

Data was presented to SCENT in two formats, typed and untyped. The typed data set contains 17 of the full 18 attributes, it excludes only the animal name from the database. The untyped data set also does not contain the animal name label but more significantly the type attribute was removed and the number of legs attribute normalised.

Feature	Label	Value
2	Hair	1
3	Feathers	0
4	Eggs	1
5	Milk	1
6	Airbourne	0
7	Aquatic	1
8	Predator	1
9	Toothed	0
10	Backbone	1
11	Breathes	1
12	Venomous	0
13	Fins	0
14	Legs	4
15	Tail	1
16	Domestic	0
17	Catsize	1

Table 1: The input vector for the platypus instance in Zoo data set.

4.1 Results

Figure 3 shows a typical tree produced using SCENT with 101 untyped vectors of the Zoo data set. Each of the data vectors is clearly categorised at both super and sub-ordinate levels. The tree structure produced using each data set was labelled in two ways. Firstly, the data vectors represented by each leaf node of the tree were labelled with the associated animal name tag. Secondly, the semantic type label of each instance was examined and the leaf node and super-ordinate clusters were classified according to each of the 7 labelled groups. Each super-ordinate class is further classified at each subsequent level of the tree. All instances of mammals are grouped together, in the leftmost clusters of the tree, and are subdivided into large-predatory, large-non-predatory, small, aquatic-legged, aquatic-finned. The middle cluster categorises fish and birds. The birds are classified in two subtrees, as predatory or non-predatory. The right hand cluster represents a variety of vectors from the less well represented classes of animal. It can therefore be seen that the emergent clusters produced by SCENT often correspond with natural taxonomic groups.

The SCENT program produces similar tree structures to that shown in Figure 3 for both the typed and untyped data sets. It is interesting to note that the absence of the type label, in the untyped data set, has little impact on the structure of the tree or its hierarchical classification. The structures of the trees produced by SCENT are

highly similar when clustering both typed and untyped data and the classification hierarchy has the same super-ordinate clusters as those identified in the type label. Typical misclassifications seen in SCENT, such as aquatic mammals (e.g., dolphin) grouped in the fish cluster and the instance of newt classified with the reptile cluster, are no more common in the untyped data than in the typed data.

The SCENT classification of the Zoo data can be compared to that produced by a hierarchical version of the ART network, HART [Bartfai 1995]. Both networks clearly become increasingly specific at lower levels and in that sense are *hierarchical*, however, there is no evidence of super-ordinate classification being developed within the HART tree structure, consequently the categorisation classes do not reflect natural groupings. The lower level classes of the SCENT model, do however often give natural sub-groupings, because of the guiding super-ordinate organisation developed during the tree's evolution.

Table 2 shows the category prototypes produced by SCENT for the mixed middle cluster of figure 3 (referred to as class B). The top level classification is almost all defined by the attributes no-hair, no-milk, has-eggs, has-backbone, has-tail. This picks out more than one taxonomic group. Sub-categories are specified more precisely. For example sub-classes B1 and B3 (representing all the instances of birds) are prototypically all of the above attributes together with feathered, toothed, breathing, not-venemous and not finned whilst B2 (representing the instances of fish) has all the above features plus not feathered, not airbourne, not breathing, not legged, aquatic, toothed and finned.

The superordinate prototypes produced by SCENT do not rigidly fix the sub-ordinate prototype, that is, a sub-ordinate category may contain a "don't care" item where its superordinate category has indicated a definite value of this attribute. This effect may arise from the stochastic process of node creation within the model, and it can provide flexibility in the classification of potentially anomalous data.

It should be noted that the prototypes developed by either the SCENT or HART models are influenced by the semantic labelling present in the data set. Anomalous labelling within data vectors must therefore be considered when viewing these results, especially against biological taxonomic classification.

The bottom-up development of hierarchical models, such as HART, forces the propagation of any mis-identification of critical features through to the developed prototypes. Models which follow a top down approach to classification, such as SCENT, do not suffer from this problem. The initial selection of critical features is made using the variance of the entire data set, thus producing a coarse grained measure of classification. Subsequent selection of critical features is made on progressively smaller sets of input vectors (as patterns are classified, fewer need classifying), the classification therefore becomes more fine grained or focused as the tree evolves.

The development of classification hierarchy models such as HART and SCENT allows for rapid high level analysis of data sets. Critical features of the data are easily identified, as are any features of little statistical significance, at a number of levels within the natural hierarchy of the data set.

The SCENT model is capable of learning stable hierarchical clusters that include both super and sub categories from semantic data.

5 Analysis and Conclusion

The SCENT model contains many elements of stochasticity, such as: the random order of vector presentation, the noisy process of growth, and the non-deterministic pruning of unsuccessful growth. Whilst this allows for exploration of the architectural and classification spaces available to the model, it also imples that each run produces a unique structure. Stability and repeatability therefore become important aspects of SCENT's performance. In order to address this issue, the two zoo data sets, with and without the type label, together with the picture data were presented for four separate runs. The structural features of the resulting trees are presented in Table 3.

It is apparent, first of all, that the there is variation in the overall structure of trees produced in different runs, however this variance is not excessive.

Overall the trees produced for the picture data set were larger than those produced for either of the zoo sets. This is accounted for exclusively by these trees having a greater branching factor; indeed the zoo data tended to produce slightly deeper trees. In view of the model's growth criterion for new clusters, downwards for dense data points and sideways for spatially separated data, this implies that the picture data has greater spatial separation.

The two zoo data sets produced trees of slightly different shape. The removal of the type label caused the trees to be shallower but with more branches, which as before shows the untyped data to be the more spatially separated. The reason for this is that two similar vectors with identical type fields are slightly less similar when the type field is removed. The full theoretical relationship between training data, in general, and the resulting structures produced by SCENT is an issue currently being investigated. From

the results and analysis presented here it can be seen that the SCENT model is capable of discovering interesting structure in unlabelled data, as well as providing a straightforward codevector reduction.

References

Bartfai, G. (1995). An ART-based Modular Architecture for Learning Hierarchical Clusterings, *Neurocomputing*, 1995.

Butchart, K., Davey, N. & Adams, R. G. (1995). A Comparative Study of Two Self-Organising and Structurally Adaptive Dynamic Neural Tree Networks. *In Proceedings of Applied Decision Technologies Conference (ADT)* 1995.

Butchart, K., Davey, N., Adams, R. (1996). Hierarchical Classification with a Stochastic Competitive Evolutionary Neural Tree, In *Proceedings of ICNN96*, Vol. 2, pp 1372 - 1377. 1996.

Butchart, K., Davey, N. & Adams, R. G. (1997). An Investigation into the performance and representations of a Stochastic Evolutionary Neural Tree. *Proceedings of the International Conference on the Applications of Neural Networks and Genetic Algorithms* (ICANNGA 97), Springer Verlag.

Gale, T. (1997). Perception and Semantic Information in Human Object Recognition: a Neuropsychological and Connectionist study. *PhD Thesis*, University of Hertfordshire, 1997

Martinetz , T., Berkovich ,S. and Schulten ,K. 1993. Neural-Gas Network for Vector Quantisation and its Application to Time-Series Prediction. *IEEE transactions on Neural Networks.* vol. 44 (4). July 1993.

Murphy, P.M. and Aha, D. W. (1992). Repository of Machine Learning Databases, Technical Report, Department of Information and Computer Science, University of California, CA, 1992.

Tables and Figures

Feature	Label	Class B	SubClass B1	SubClass B2	SubClass B3	SubClass B4	SubClass B5
2	Hair	no	no	no	no	*	no
3	Feathers	*	yes	no	yes	*	no
4	Eggs	yes	yes	yes	yes	yes	*
5	Milk	no	no	no	no	no	no
6	Airbourne	*	*	no	*	*	no
7	Aquatic	*	*	yes	*	*	*
8	Predator	*	no	*	yes	*	yes
9	Toothed	*	no	yes	no	*	yes
10	Backbone	yes	yes	yes	yes	*	yes
11	Breathes	*	yes	no	yes	*	*
12	Venomous	*	no	*	no	*	*
13	Fins	*	no	yes	no	no	*
14	Legs	*	*	no	*	*	*
15	Tail	yes	yes	yes	yes	*	yes
16	Domestic	*	*	*	no	*	no
17	Catsize	*	*	*	*	*	no

Table 2: Catgory Prototypes for middle cluster of figure 3, the Zoo Tree. “Yes” indicates attribute presence (weight value >=1.0). “No” indicates absence of attribute (weight value <=0.0). “*” indicates a don’t care value for the attribute (1.0> weight value >0.0)

Data Source	Leaf Nodes	Av. Branching	Av. Depth
Pictures	30	3.2	3
	38	3.5	3.2
	34	3.2	3.2
	27	3.2	2.7
	Average 32	**Average 3.27**	**Average 3.02**
	St. Dev. 4.1	**St. Dev. 0.12**	**St. Dev. 0.20**
Zoo Typed			
	30	2.53	3.57
	23	3.30	2.65
	24	2.77	3.58
	30	2.56	3.77
	Average 27	**Average 2.79**	**Average 3.39**
	St. Dev. 3.3	**St. Dev. 0.31**	**St. Dev. 0.44**
Zoo Untyped			
	29	3.00	3.03
	32	3.07	3.25
	32	2.94	3.00
	24	3.18	2.83
	Average 29	**Average 3.05**	**Average 3.02**
	St. Dev. 3.3	**St. Dev. 0.09**	**St. Dev. 0.15**

Table 3: Summary of the tree structures produced by SCENT over the three data sets and four runs.

Figure 1: The 51 high contrast pictures from the 153 pictures in the complete data set