Visual Sensitivity Analysis for Artificial Neural Networks*

Roberto Therón, Juan Francisco De Paz

Departamento de Informática y Automática
Facultad de Ciencias - Universidad de Salamanca
Plaza de la Merced s/n. 37008. Salamanca (Spain)
theron@usal.es, fcofds@usal.es

Abstract. A challenge in ANN research is how to reduce the number of inputs to the model in high dimensional problems, so it can be efficiently applied. The ANNs black-box operation makes not possible to explain the relationships between features and inputs. Some numerical methods, such as sensitivity analysis, try to fight this problem. In this paper, we combine a sensitivity analysis with a linked multi-dimensional visualization that takes advantage of user interaction, providing and efficient way to analyze and asses both the dimension reduction results and the ANN behavior.

1 Introduction

Many disciplines (such as bioinformatics, economics, climatology, etc.) face a classification or prediction problem involving large number of features. However, the high dimensionality of the data can lead to inaccurate results or even disqualify the use of machine learning methods. The *curse of dimensionality* stipulates that it is hard to apply a statistical technique to high-dimensional data.

Feature selection and dimension reduction techniques are both used to remove features that do not provide significant incremental information. Numerous studies have revealed that in high-dimensional data, feature selection and dimension reduction methods are essential to improve the performance of a classifier report on dimension reduction techniques such as principal component analysis (PCA) or factor analysis [1].

Despite the great success in many fields, ANNs are still regarded as black-box methods [2] where it is difficult for the user to understand the nature of the internal representations generated by the network in order to respond to a certain problem. In order to overcome this problem, different rule extraction and numerical methods are applied to study the contribution of variables in a neural network; sensitivity analysis is one of the most broadly used [3].

In recent years the field of information visualization has played an important role providing insight through visual representations combined with interaction techniques that take advantage of the human eye's broad bandwidth pathway

* This work was supported by the MCyT of Spain under Integrated Action (Spain-France) HF2004-0277 and by the Junta de Castilla y León under project SA042/02.

to the mind, allowing experts to see, explore, and understand large amounts of information at once. In this work we combine a sensitivity analysis with a linked multi-dimensional visualization, mainly based on interactive parallel coordinates[4], that can help to understand the behavior of the neural network, analyze its sensitivity, and provide a way to interpret the relationship between features and outputs.

1.1 Related work

Tzeng and Ma [5] provide a survey of several visualization techniques for understanding the learning and decision-making processes of neural networks. These techniques include Hinton diagrams, bond diagrams, response-function plots, hyperplane diagrams, and trajectory diagrams [6], that are used to illustrate the idea of neural networks but are not practical due to the difficulty of showing a large network clearly. The visualization method in [5] allows the user to probe into the data domain and visualize the corresponding network, errors, and uncertainty visualization, by means of Parallel coordinates [4], to help both the designer and the user of a neural network. In [7] an interactive visualization tool for feed-forward neural networks, based on tree/graph visualization, is described. Although the visualization tool is useful both as an educational device (to aid in the understanding of neural networks, search spaces, and genetic drift), and as a practical tool for solving complex problems with neural networks, the authors recognize that its main limitation is that the graphical feedforward network depiction does not scale well to networks with large numbers of nodes.

A projection on a lattice of hypercube nodes to visualize the hidden and output node activities in a high dimensional space is used in [8]. Scatterograms of the images of training set vectors in the hidden space help to evaluate the quality of neural network mappings and understand internal representations created by the hidden layers. Visualization of these representations leads to interesting conclusions about optimal architectures and training of such networks.

In [9], Cook and Yin discuss visualization methods for discriminant analysis adapted from results in dimension reduction for regression (sliced inverse regression and sliced average variance estimation). The method are good identifying outliers. The graphical representations used for regression visualization are Summary plots, where the structural dimension of the data is used, so such plots have the minimal complexity needed to explain the structure of the model and to make predictions.

2 Visual sensitivity analysis

In order to add power to ANNs in their explanatory capacity and understand the complex relationships that occur between the variables, sensitivity analysis [10][11][3] have been used. Having trained a neural network, an input sensitivity analysis is conducted on the trained network, using the training data.

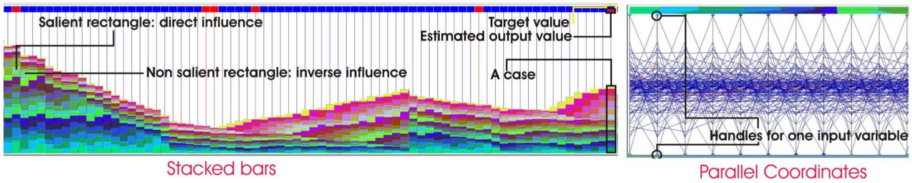

Fig. 1. Visual techniques used in visual sensitivity analysis

In the Jacobian matrix S, each line represents an output to the network and each column represents an input to the network, so that the element S_{ki} represents the sensitivity of the output y_k with respect to the input x_i, calculated as a partial derivative of the output with respect to the input, $S_{ki} = \frac{\partial y_k}{\partial x_i}$. This way, the higher the value of S_{ki}, the more important it is x_i with respect to y_k. The sign indicates the kind of influence (direct or inverse).

From a practical point of view it is more interesting to understand how different inputs affect to a given output for the training pairs (cases). Thus, the purpose of the visual sensitivity analysis is to provide a representation of the relationships between the output y_k with each of the inputs x_i for each of the cases. Furthermore, the inputs and their output values will also be represented to be able to compare the input data and the sensitivity analysis.

2.1 Visualization techniques

In this section we will explain the information visualization techniques used in the visual sensitivity analysis. The interface consists of two clearly differentiated areas: one for stacked bars and another one for parallel coordinates. Both areas are linked so the interaction on one implies changes in the other one.

Each case in a problem is represented as a stacked bar (see figure 1), divided into as many fragments (rectangles) as input variables. On top of each bar, both the output value estimated by the ANN and the target are represented.

Each stacked bar is color coded in order to distinguish each of the variables of the case. The height of the rectangle is used to represent the value of the variable. On the other hand, bars are represented in 3D, so that the rectangles with salient appearance represent positive values (direct influence), while negative values (inverse influence), are represented without relief (see figure 1).

Target and estimated output values are also color coded, from blue (lowest value, 0) to red (highest value, 1). This way, we can determine in a visual and quick way erroneous network estimates (note the last case on the right in figure 1) or **see** the group to which each case belongs.

Parallel Coordinates Plot (PCP) [4] is one of the most successful techniques in multi-dimensional visualization. Each case is drawn as a polyline passing through parallel axis, which represent the input variables (see figure 2).

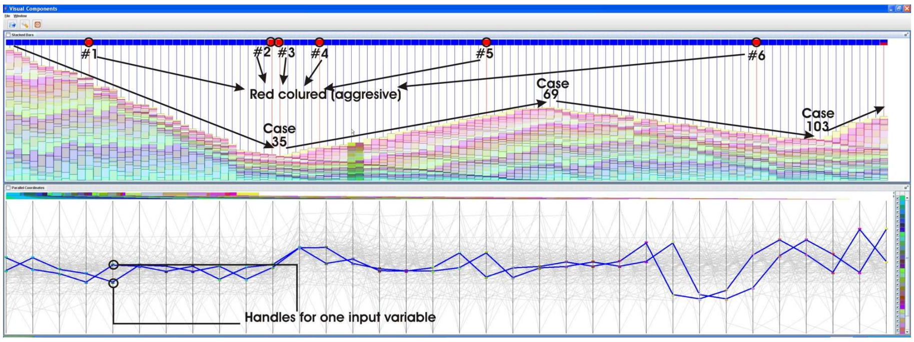

Fig. 2. Selecting cases: bars and polylines

Each axis represents an input variable; thus, the handles (see figure 1) are colored using the corresponding color of the rectangles in the stacked bars. A case (bar) is also a polyline in the PCP (see two cases highlighted in figure 2).

Our aim was to design a highly interactive visual interface that would allow to compare relationships between the variables and their values, through the cases and to determine the ranges of variable values that separate the individual cases into groups. Several interaction techniques [12] have been integrated to allow brushing [13], linking, focus + context, etc., for exploratory analysis and knowledge discovery purposes. Thus, it is possible to select one or several bars and the corresponding polylines are highlighted, and vice versa (see figure 2); the order of bars and axis can be altered; tooltips are used to give details on demand; handles in axis can be used to filter cases based upon interesting variables ranges (see how the handles were used to filter cases in figure 2), etc.

3 Case study: Aggressive behavior in high school students

Following, the visual sensitivity analysis for the aggressive behavior in high school students is explained. 111 students of 7 schools with ages ranging from 14 to 17 years answered to 114 questions. The dimension of the problem was reasonably high for building a neural network classifier, so a factor analyis with a PCA extraction method was performed to reduce the number of variables. As a result, 34 factors were extracted, i.e., the actual number of variables used to train a Multilayer Perceptron (MLP). Having the input/output and sensitivity data, this visualization technique can be used with other types of ANNs.

Once the MLP was trained, an analysis of sensitivity was carried out. The results of this process are then used in the visual sensitivity analysis, in order to study the relationship of the input variables with the aggressive behavior of students (i.e., the target used for the network training).

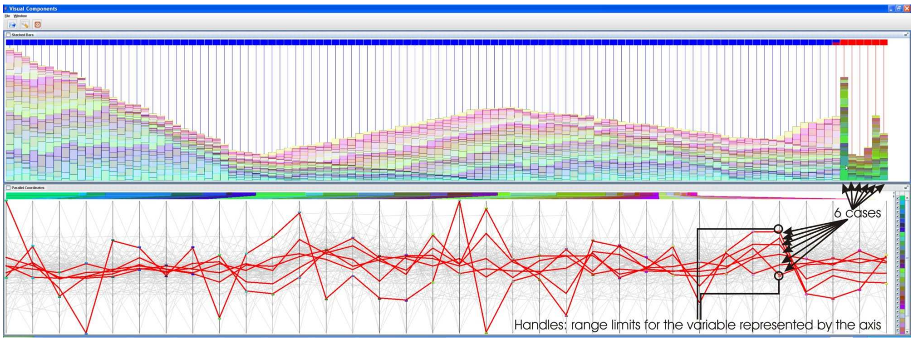

Fig. 3. Visualization of sensitivity ranges for aggressive behavior

3.1 Sensitivity analysis

The visual sensitivity analysis for the MLP trained after dimension reduction can be seen in figure 2. Red colors on outputs (on top of stacked bars) represent an aggressive behavior. Note an estimate error on the last case: the target (above) is blue (non aggressive), while the estimated value (below) is red (aggressive).

Focusing in the stacked bars, different areas can be observed, depending on how different cases (students) were affected by the components. Concretely, in a first group, a falling trend in the influence of the inputs to the output is found; this can be observed up to the third aggressive student (red colored target, case 35). Then a growing trend begins arriving to a local maximum, a non aggressive student (case 69). Finally, there is another falling phase arriving to student 103 and a small growing one up to the 111.

This result is quite curious; we had 34 neurons in the input layer and the changes take place every 34+1 cases. In order to explain this, two neighbor bars were selected and the polylines examined: they are almost identical but displaced (see figure 2). All couples of neighbor bars (cases) offer the same result, except for trend change places. That is, during the training, the influence of the input neurons in the output goes moving toward the end of the input neurons, then a small variation takes place. The training order does not change this situation.

An interesting question is if there exists a value for the coordinates that separate the aggressive students. The cases were ordered according to an increasing value of target. Then, the aggressive cases were selected so the ranges were automatically delimitated by the axis handles (see figure 3). The only polylines that were active and highlighted were those corresponding to the selected bars (aggressive students). The remaining cases are drawn with soft colors in the background so the context of the problem is not lost.

3.2 Target data

This analysis is similar to the previous one and complements it. Now, instead of the influences, the the MLP input values are represented. The result is shown in

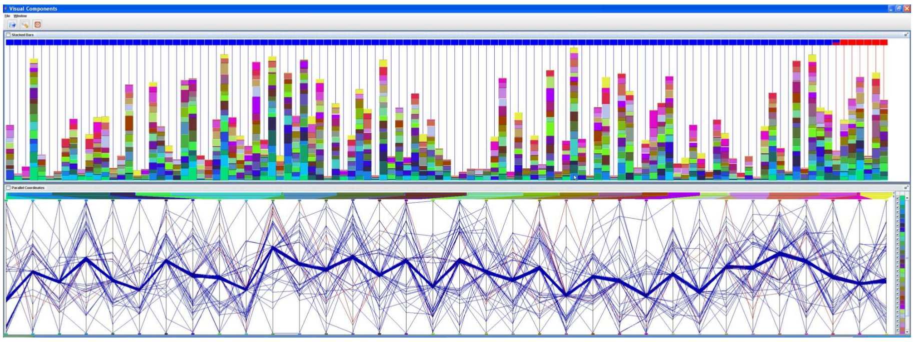

Fig. 4. Visual inputs analysis

figure 4. It can be observed that the height of the bar does not follow any certain pattern: no relationship similar to that observed for the analysis of sensitivity can be found. In this case, few conclusions can be reached starting from the stacked bars. There are many cases in which the values of the answers are low (very small bars), but this does not separate the aggressive students. In this case, the color code of the bars, as opposed to what happened in the sensitivity analysis, does not contribute much information.

In this case, it is more useful the parallel coordinates plot. It can be easily seen that a mass of polylines exists in the central part of the plot. They correspond to the bars with the smaller heights. This permits to understand that, possibly, there was a group of students that were not interested in the survey and they answered similarly to all the questions. Another explanation is that they may form a differentiated class; this should be kept in mind when choosing cases for the sample. Repeating the previous process, ordering the bars according to height, and selecting only the smaller bars, it can be seen that these students actually form a separated group (see the polyline pattern in figure 5).

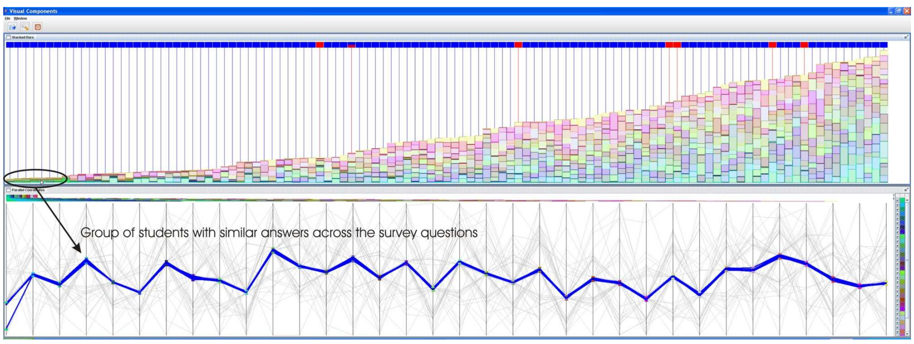

Fig. 5. Discovering a group of students

Now, the main question is if it is possible to determine if a student is aggressive according to his/her answers. By selecting the aggressive students (as context, the rest are maintained in the background). The handles in the parallel axis are automatically placed so they indicate that if the answers are inside those ranges, the student is aggressive (see figure 6). This a quick and easy way to determine if the variables are actually good to classify the students. In this particular situation the result has been affirmative. Note how aggressive students form an independent group: there are not blue polylines selected anymore.

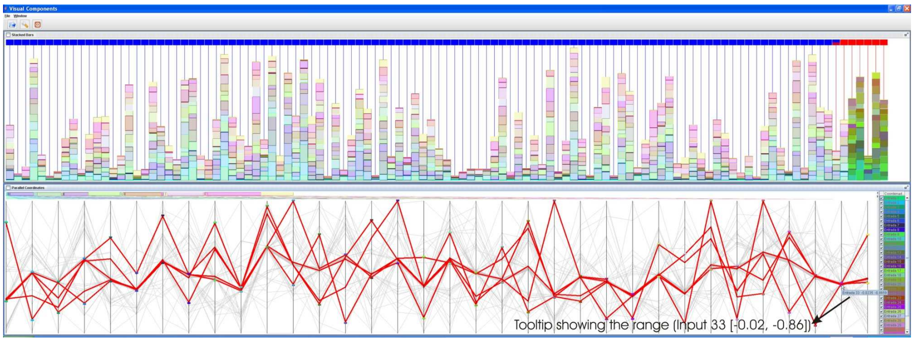

Fig. 6. Visualization of sensitivity ranges for aggressive behavior

Finding the smaller range in the axis will provide the more important inputs to classify the students behavior. Note the tooltip of input 33 (figure 6) showing the range in which the answer of an aggressive student should be. Remember that inputs are the result of dimensionality reduction, so a conversion to the actual answers of the student should be performed. Furthermore, we can go on removing axis (inputs) and seeing that there are not aggressive cases inside the range. After this process the most important variables that allow the isolation of the students are inputs 1, 23, 25, 26, 28, 32 and 34 (see checked boxes on the right of the PCP, figure 7). Same behavoir is observed in the sensitivity analysis.

4 Conclusions

A novel method for the visualization and exploration of the relationship between features and outputs in ANNs was presented. The combination of sensitivity analysis with information visualization techniques for multi-dimensional data provides a solution to face the curse of dimensionality. As case study, a visual sensitivity analysis for a MLP classifier of aggressive students has been shown. Although PCPs and Stacked bars are valid for a high number of variables, future work will be testing the proposed technique limitations and if these can be faced with other information visualization techniques.

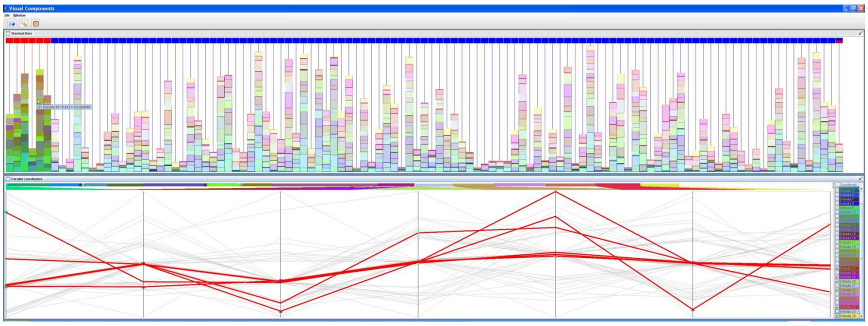

Fig. 7. Discovering the relevant factors in input analysis

References

1. Berrar, D.P., Downes, C.S., Dubitzky, W.: Multiclass cancer classification using gene expression profiling and probabilistic neural. In: Pacific Symposium on Biocomputing. Volume 8. (2003) 5–16
2. Sjoberg, J., Zhang, Q., Ljung, L., Benveniste, A., Delyon, B., Glorennec, P., Hjalmarsson, H., Juditsky, A.: Nonlinear black-box modeling in system identification: a unified overview. Automatica **31** (1995) 1691–1724
3. Gevrey, M., Dimopoulos, I., Lek, S.: Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecological Modelling, **160** (2003) 249–264
4. Inselberg, A.: The plane with parallel coordinates. The Visual Computer **1** (1985) 69–91
5. Tzeng, F.Y., Ma, K.L.: Opening the black box - data driven visualization of neural networks. In: Proceedings of IEEE Visualization '05 Conference. (2005) 383–390
6. Craven, M., Shavlik, J.: Visualizing learning and computation an artificial neural networks. International Journal on Artificial Intelligence Tools **1** (1992) 399–425
7. Streeter, M.J., Ward, M.O., Alvarez, S.A.: Nvis: An interactive visualization tool for neural networks. In: Proceedings of SPIE Symposium on Visual Data Exploration and Analysis. (2001) 234–241
8. Duch, W.: Visualization of hidden node activity in neural networks: I and ii. In: Proceedings of the International Conference on Artificial Intelligence and Soft Computing. (2004) 38–49
9. Cook, R.D., Yin, X.: Special invited paper: Dimension reduction and visualization in discriminant analysis (with discussion). Australian & New Zealand Journal of Statistics **43** (2001) 147–199
10. Hwang, J.N., Choi, J.J., Oh, S., II, R.J.M.: Query-based learning applied to partially trained multilayer perceptrons. IEEE Transactions on Neural Networks **2** (1991) 131–136
11. Fu, L., Chen, T.: Sensitivity analysis for input vector in multilayer feedforward networks. In: Proceedingd of IEEE International Conference on Neural Networks. Volume 1. (1993) 215–218
12. Keim, D.A.: Information visualization and visual data mining. IEEE Transactions on Visualization and Computer Graphics **8** (2002) 1–8
13. Becker, R.A., Cleveland, W.S.: Brushing scatterplots. Technometrics **29** (1987) 127–142