8th Seminar on Neural Network Applications in Electrical Engineering, *NEUREL-2006*
Faculty of Electrical Engineering, University of Belgrade, Serbia, September 25-27, 2006
http://neurel.etf.bg.ac.yu, http://www.ewh.ieee.org/reg/8/conferences.html

Data Mining a Prostate Cancer Dataset Using Neural Networks

Kenneth Revett

Abstract — **Prostate cancer remains one of the leading causes of cancer death worldwide, with a reported incidence rate of 650,000 cases per annum worldwide. The causal factors of prostate cancer still remain to be determined. In this paper, we investigate a medical dataset containing clinical information on 502 prostate cancer patients using the machine learning technique of rough sets and radial basis function neural network.. Our preliminary results yield a classification accuracy of 90%, with high sensitivity and specificity (both at approximately 91%). Our results yield a predictive positive value (PPN) of 81% and a predictive negative value (PNV) of 95%.**

Keywords — **cancer classifier ,machine learning, prostate cancer dataset, reducts, Rough sets**

I. Introduction

Prostate cancer is the second leading cause of mortality in men, exceeded only by lung cancer. The cause(s) of this form of cancer remain to be elucidated, but factors such as diet, heredity, and environmental factors that effect male hormones (androgens) have been implicated in epidemiological studies [1-3]. Currently, two standard tests are used for early detection of prostate cancer:

- Digital rectal examination (DRE). With the DRE, a physician palpates the prostate in order to feel lumps or masses.
- PSA test. The PSA blood test measures the level of a protein called prostate-specific antigen. It is able to detect early prostate cancer, although it has limitations.

There are many unresolved questions surrounding PSA testing. The test is not accurate enough to either completely rule out or confirm the presence of cancer. Current treatments entail chemotherapy, surgery or a combination of the two depending on the stage of disease progression. Further, the incidence of prostate cancer increases with age. This will present an increased incidence as the world population tends towards increased longevity. This trend is alarming and warrants investigating the causative factors in prostate cancer through all available means. In this paper, we present the results of a machine learning technique based on rough sets and neural networks (RBF) to the study of a clinically relevant prostate cancer dataset.

In this study, we investigate a dataset containing data on 502 (29%/71% live/dead) patients that were diagnosed with prostate cancer. The dataset contains 18 attributes including the decision attribute (see section 2.1 for a listing of the attributes) with 27 missing values (0.3%). We investigated this dataset with respect to the following: i) attribute pruning, ii) classification accuracy and iii) rule induction. Pruning (dimensionality reduction) removes variables that are not directly related to the classification process. This feature of rough sets makes the dataset much easier to work with and may help to highlight the relevant classification features of the data. Once the redundant features have been pruned from the dataset, a RBF neural network was developed and applied to the data to perform the actual classification. In the next section, we present a very brief overview of rough sets, followed by the application of a Radial basis Function (RBF) neural network, followed by a results section and lastly a summary of this work.

II. Methods

In this study, we employed the rough sets paradigm as a pre-processing step. The purpose of this step was to reduce the dataset in terms of attributes as much as possible in order to minimise the computational work required to train a neural network based classifier. We next briefly describe the rough sets algorithm and also the neural network classifier – the radial basis function neural network (RBF). We use the classification accuracy of rough sets as a benchmark for the classification accuracy obtained from the RBF algorithm.

Kenneth Revett is with the University of Westminster, Harrow School of Computer Science, London, England HA1 3TP, (phone +44(0)207-911-5000, fax +44(0)207-911-5906, email: revettk@westminster.ac.uk).

A. *Rough Sets*

In order to apply rough sets, which is a supervised machine learning paradigm, the data set must be transformed into a decision table (DT) from which rules are generated to provide an automated classification capacity. In generating the decision table, each row consists of an observation (also called an object) and each column is an attribute, with the last one as the decision for this object {d}. Formally, a DT is a pair A = (U, A$\cup${d}) where d ϖ A is the *decision attribute,* where U is a finite non-empty set of objects called the *universe* and A is a finite non-empty set of attributes such that a:U->V_a is called the value set of a. Rough sets seeks data reduction through the concept of equivalence classes (through the indiscernibiliy relation). By generating such classes, one can reduce the number of attributes in the decision table by selecting any member of the equivalence class as a representation of the entire class. This process generates a series of *reducts* – which are subsequently used in the classification process. Finding the reducts is an NP-hard problem, but fortunately there are good heuristics that can compute a sufficient amount of approximate reducts in reasonable time to be usable. An order based genetic algorithm (o-GA) (cf [4,5]) is used to search through the decision table for approximate reducts which result in a series of 'if..then..' decision rules. We then apply these decision rules to the test data and measure specificity and sensitivity of the resulting classifications. Lastly, we examined in a systematic fashion, which attributes were required for the decision process – this can be determined by examining the correlation, coverage, and support of the attributes in the final set of decision rules [6],[7]. This provides both a classification accuracy that can be used as a benchmark for other techniques, and also a reduced decision table. This reduced DT is used as the input for training a RBF neural network and the classification results are compared with those produced by rough sets alone.

B. *Radial Basis Function Neural network*

Radial Basis Functions are powerful techniques for interpolation in multidimensional space. A RBF is a function which has built into a distance criterion with respect to a centre. Radial basis functions have been applied in the area of neural networks where they may be used as a replacement for the sigmoidal hidden layer transfer function in multilayer perceptrons. RBF networks have 2 layers of processing: In the first, input is mapped onto each RBF in the 'hidden' layer. The RBF chosen is usually a Gaussian. In regression problems the output layer is then a linear combination of hidden layer values representing mean predicted output. The interpretation of this output layer value is the same as a regression model in statistics. In classification problems the output layer is typically a sigmoid function of a linear combination of hidden layer values, representing a posterior probability. Performance in both cases is often improved by shrinkage techniques, known as ridge regression in classical statistics and known to correspond to a prior belief in small parameter values (and therefore smooth output functions) in a Bayesian framework.

RBF networks have the advantage of not suffering from local minima in the same way as multilayer perceptrons. This is because the only parameters that are adjusted in the learning process are the linear mapping from hidden layer to output layer. Linearity ensures that the error surface is quadratic and therefore has a single easily found minimum. In regression problems this can be found in one matrix operation. In classification problems the fixed non-linearity introduced by the sigmoid output function is most efficiently dealt with using iterated reweighted least squares.

RBF networks have the disadvantage of requiring good coverage of the input space by radial basis functions. RBF centres are determined with reference to the distribution of the input data, but without reference to the prediction task. As a result, representational resources may be wasted on areas of the input space that are irrelevant to the learning task. A common solution is to associate each data point with its own centre, although this can make the linear system to be solved in the final layer rather large, and requires shrinkage techniques to avoid overfitting. For a more complete description, please see [8].

III. Results

In this study, we explored the dataset from 2 different perspectives. Initially, we wished to see how rough sets could be used to reduce the dimensionality of the data – in addition to providing a classification accuracy benchmark. We then employed a powerful RBF neural network technique. In addition to pure classification accuracy, we wished to determine if data reduction using rough sets enhanced the classification accuracy of the RBF.

Table 1. The dataset used in this study. The numbers in parentheses refer to the number of categories for that particular attribute. For details on category values, see reference [9].

Attribute Name	Attribute Type
Patient number	Double
Stage	Double
Treatment	Integer (4)
Dtime(follow up time in months)	Double
Date on Study	Double
Age	Integer
Weight index (wt(kg) – ht(cm) + 200)	Integer
Pf	Integer (4)
Hx (history of cardio-vascular disease)	Double
Sbp Systolic bp	Double
Dbp Diastolic bp	Double
EKG	Integer (7)
Hg (serum haemoglobin (g/100ml)	Double
Sz (size of primary tumour (cm^2)	Double
Index of Stage and Histology	Double
Serum Prostatic Acid Phosphatase	Double
Bone Metatastases	Double
Status	Double (10)

Table 2. Confusion matrices from a set of five randomly selected classification tasks on the test case (using 70/30 train/test) from 20 randomly selected tests

Test1	**Alive**	**Dead**	
Alive	**34**	**3**	0.918919
Dead	7	**81**	0.920455
	0.829268	0.964286	**0.92**
Test2			
Alive	**28**	**3**	0.903226
Dead	**11**	**83**	0.882979
	0.717949	0.965116	**0.888**
Test3			
Alive	**39**	7	0.847826
Dead	**9**	**70**	0.886076
	0.8125	0.909091	**0.872**
Test4			
Alive	**37**	**4**	0.902439
Dead	**10**	**74**	0.880952
	0.787234	0.948718	**0.888**
Test5			
Alive	**35**	**4**	0.897436
Dead	**15**	**71**	0.825581
	0.7	0.946667	**0.848**

Table 3. Classification accuracy of the RB F neural network with and without pre-processing using the rough sets algorithm. The first row is *without* dimensionality reduction (pre-processing) and the second row is *with* dimensionality reduction.

Training	testing
91.0%	90.2%
95.3%	94.1%

IV. Discussion

The results from this study indicate that the combination of rough sets and the RBF neural network provide an enhanced classification accuracy with this dataset. This could be the result of removing redundancy and conflicting information from the decision table. To the

authors knowledge, there has not been any previous study which investigated the combination of rough sets and RBF neural networks on this type of dataset.

The combination provides a robust method for reducing non-informative attributes. The resulting classifier was enhanced when the attribute reduction was applied, indicating that there was indeed contradictory data within the decision table that prevented the RBF neural network from escaping some local minimum(a) within the search space. With sparse and important datasets that are derived from the biomedical literature, one must take every step necessary to ensure that the maximal amount of information has been extracted in the examination of such datasets. Rough sets provides a unique method for performing this operation in a standardised method. Inherent to its operation is the concept of redundancy excavation – through the concepts of equivalence classes, attributes that are not informative can be removed without resulting in a drop in the classification accuracy. This ability is an extremely valuable tool for datamining. In cases where the classification accuracy of rough is sub-maximal, this usually indicates that the data contains inconsistencies. For instance, the same set of antecedents may yield two different decision class values. This issue was not addressed in this paper – although through the concept of approximate reducts and variable precision reducts, inconsistencies can be handled quite reasonably. In the current dataset, the classification accuracy was higher with the RBF – which in the end justifies the combination approach employed in this preliminary study. It is quite possible that the non-linearity of the hidden layer in the RBF was able to handle
possible inconsistencies within the dataset. This combination of machine learning algorithms provides data miners with the tools required to extract useful and important information from small biomedical datasets. In the future, we will expand the set of classifiers and also explore approximate reducts in order to determine the extent of the inconsistencies within the data and also as a means of removing them.

V. REFERENCES

[1] D.F. Andrews and A.M. Herzberg, Data, a Collection of Problems from Many Fields for the Student and Research Worker. Springer-Verlag, New York, 1985.

[2] A.Jemal, T. Murray, E. Ward E, A. Samuels, Cancer Statistics, 2005. CA Cancer J Clin 55:10-30, 2005.

[3] P. Kalra, J. Togami, G. Bansal, A.W.Partin , M.K. Brawer, R.J. Babaian, L.S. Ross, & C.S. Niederberger: A Neurocomputational Model for Prostate Carcinoma Detection: Cancer, Volume 98, Issue 9 ,November 2003.

[4] H.S. Nguyen & A. Skowron, Quantization of real-valued attributes, Proc Second International Conference on Information Science, pp 34-37, 1995.

[5] Øhrn,. "Discernibility and Rough Sets in Medicine" Tools and Applications. Department of Computer and Information Science. Trondheim, Norway, Norwegian University of Science and Technology: 239, 1999.

[6] [Z. Pawlak, Rough Sets, International Journal of Computer and Information Sciences, 11, pp. 341-356, 1982

[7] J. Wroblewski.: Theoretical Foundations of Order-Based Genetic Algorithms. Fundamenta Informaticae 28(3-4) pp. 423–430, 1996.

[8] Haykin, S. Adaptive Filter Theory, 3rd ed., Upper Saddle River, NJ: Prentice-Hall, 1996

[9] DP, Green SB: *Bulletin Cancer*, Paris 67:477-488, 1980.