A heuristic approach in hepatic cancer diagnosis using a probabilistic neural network-based model

Marina Gorunescu[1], Florin Gorunescu[2], Marius Ene[2], and Elia El-Darzi[3]

[1] Faculty of Mathematics and Computer Science
University of Craiova,
13 A.I. Cuza, Craiova, Romania
(e-mail: gorun@euroweb.ro)
[2] Department of Mathematics, Biostatistics and Computer Science
University of Medicine and Pharmacy of Craiova,
2-4 Petru Rares, Craiova, Romania
(e-mail: gorun@umfcv.ro, enem@umfcv.ro)
[3] Harrow School of Computer Science
University of Westminster, London
Watford Road, Northwick Park, Harrow HA1 3TP, UK
(e-mail: eldarze@wmin.ac.uk)

Abstract. This paper is focusing on the application of a probabilistic neural network-based model in diagnosing hepatic diseases. In the diagnose process, the physicians compare numerical medical data against prior knowledge in order to determine the right diagnostic. Neural networks are ideal in recognizing diseases using representative examples since there is no need to provide a specific algorithm on how to identify the disease. The goal of this paper is to explore a PNN-based approach to determine the (near) optimum diagnosis for hepatic cancer. As concerns the concrete program, a Java implementation is provided as well.
Keywords: probabilistic neural networks, Monte Carlo approach, hepatic diseases diagnosis, Java implementation.

1 Introduction

Hepatocellular carcinoma (HCC), briefly hepatic cancer, represents a primary malignant tumor of the liver that ranks fifth in frequency among all malignancies in the world. HCC is increasing in many countries, especially in areas where hepatitis C virus (HCV) infection is more common than hepatitis B virus (HBV) infection. The diagnosis of HCC is difficult in the early stages, most of the patients being diagnosed in advanced stages. Although alpha-fetoprotein (AFP) is the most important tumor marker for the diagnosis of HCC, a considerable proportion of HCC's do not produce AFP, making early diagnosis difficult with this marker alone. Imaging modalities (power Doppler, harmonic imaging, pulse inversion, etc.), combined with micro bubble contrast agents and a better understanding of the importance of serum enzymes significantly improved the rate of detection for early (small)

HCCs'. Among these detection factors, the serum enzymes analysis is by far the fastest and simplest method, representing the first step in hepatic cancer diagnosis.

The probabilistic neural network (PNN) was developed by [Specht, 1988] [Specht, 1990]. This particular type of artificial neural networks (ANNs) provides a general solution to pattern classification problems by following the probabilistic approach based on the Bayes formula. The Bayes decision theory, emerged from his celebrated formula and developed in the 1950's, takes into account the relative likelihood of events and uses a priori information to improve prediction. The network paradigm uses the Parzen estimators to obtain the corresponding probability density functions (p.d.f.) to the classification categories. In his classic paper, Parzen [Parzen, 1962] showed that a class of p.d.f. estimators asymptotically approach the underlying density function, provided that it is continuous. Cacoulos [Cacoulos, 1966] extended Parzen's method to the multivariate case. PNN uses a supervised training set to develop probability density functions within a pattern layer. Training of a PNN is much simpler than other ANNs techniques. Key advantages of PNN are that training requires only an unique pass and that the decision hipersurfaces are guaranted to approach the Bayes-optimal decision boundaries as the number of training samples grows. On the other hand, the main criticism of PNN is that all training samples must be stored and used in classifying new patterns (very rapid increase in memory and computing time when the dimension of the input vector and the training set size increase). However, to reduce the computational cost, dimensionality reduction and clustering methods are usually applied, previous to the PNN construction.

ANNs in general and PNNs especially are currently a main research area in health care modelling and it is believed that they will receive extensive application to biomedical systems in the next years ([Lin *et al.*, 2002], [Norton *et al.*, 2001], [Taktak *et al.*, 2004]). Neural networks learn by examples so the details of how to recognize the disease is not needed. We only need a set of examples (patterns) that are representative of all the variations of the specific disease. To obtain a high accuracy level in the disease recognition the patterns generally need to be selected carefully.

2 Bayes decision rule for PNNs

Bayesian decision theory is a fundamental statistical approach to the problem of pattern classification. To illustrate the formalism of the Bayes decision rule, consider the sample space Ω and $B_1, B_2, ...B_n$ a partition of Ω. Then the celebrated reverend Bayes formula (1763) is given by:

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)} \tag{1}$$

Usually, the Bayes formula becomes:

$$Posterior = \frac{likelihood \ \times \ prior}{evidence} \tag{2}$$

where $P(B_i|A)$ is known as *Posterior*, $P(B_i)$ -the *prior* probabilities, $P(A|B_i)$ -the *likelihood*, $P(A)$ -the *evidence*.

Formally, the Bayes decision rule in a simplified form is given by:

- Decision D_k: "State of nature is B_k";
- Given measurement x if the decision is D_k then the error is $P(error|x) = 1 - P(B_k|x)$;
- Minimize the probability error;
- Bayes decision rule: "*Decide* D_k *if* $P(B_k|x) > P(B_j|x), \ \forall j \neq k$" or, equivalently, "*Decide* D_k *if* $P(x|B_k)P(B_k) > P(x|B_j)P(B_j), \ \forall j \neq k$"

To illustrate the way the Bayes decision rule is applied to PNNs, consider the general case of the q-category classification problem, in which the states of nature will be denoted by $\Omega_1, \Omega_2, ..., \Omega_q$. The goal is to determine the class (category) membership of a multivariate sample data (i.e. a p-dimensional random vector $\mathbf{x}$) into one of the q possible groups $\Omega_1, \Omega_2, ..., \Omega_q$, that is, we have to make the decision $D(x) = \Omega_i, \ i = 1, \ 2, ..., \ q$, where $\mathbf{x}$ represents the sample (data vector). If we know the (multivariate) probability density functions $f_1(x), \ f_2(x), ..., \ f_q(x)$, associated with the categories $\Omega_1, \Omega_2, ..., \Omega_q$, the *a priori* probabilities $h_i = P(\Omega_i)$ of occurrence of patterns from categories Ω_i and the *loss* (or *cost*) parameters l_i associated with all incorrect decisions given $\Omega = \Omega_i$, then, according to the Bayes decision rule, we classify $\mathbf{x}$ into the category Ω_i if the following inequality holds true:

$$l_i h_i f_i(x) > l_j h_j f_j(x), \ i \neq j. \tag{3}$$

The boundaries between every two decision classes Ω_i and Ω_j, $i \neq j$, are given by the hypersurfaces:

$$l_i h_i f_i(x) = l_j h_j f_j(x), \ i \neq j, \tag{4}$$

and the accuracy of the decision depends on the accuracy of estimating the corresponding p.d.f's.

The key to using the Bayes decision rule to PNNs is represented by the technique chosen to estimate the p.d.f's $f_i(x)$ corresponding to each decision class Ω_i, based upon the training patterns set. The classical approach uses a sum of small multivariate Gaussian distributions, centered at each training sample, that is:

$$f_i(x) = \frac{1}{\sigma_i^p (2\pi)^{p/2}} \cdot \frac{1}{m_i} \cdot \sum_{j=1}^{m_i} exp\left(-\frac{\|x - x_j\|^2}{2\sigma_i^2} \right), \ i = 1, \ 2, ..., \ q, \tag{5}$$

where m_i is the total number of training patterns in Ω_i, x_j is the j-th training pattern from category Ω_i, p is the input space dimension and σ is an adjustable "*smoothing*" parameter using the training procedure. The main issue in PNNs methodology is represented by the way to determine the value of σ, since this parameter needs to be estimated to cause reasonable amount of overlap. Commonly, the smoothing factor is chosen heuristically. If σ is too large or too small the corresponding probability density functions will lead to the increase in misclassification rate. Fortunately, PNNs are not too sensitive to the very precise choice of the smoothing factor.

3 Modified Specht algorithm (Monte Carlo approach)

The only control parameter that needs to be selected for probabilistic neural network training is the radial deviation of the Gaussian densities -the smoothing factor. This section deals with one of the simplest but most robust algorithm, straight related to the Parzen-Cacoulos window classifiers, using the sum of training patterns that are classified in the right way as cost function and the Monte Carlo method for searching for the best solution. Among other statistical or Artificial Intelligence techniques, the Monte Carlo method allows us to obtain the optimization of the smoothing factor for each category with a good accuracy and saving computational effort.

Algorithm (training)

Input. Consider q decision classes $\Omega_1, \Omega_2, ..., \Omega_q$, each decision class Ω_i containing a number of m_i training patterns.

i) For each class Ω_i, $i = 1, 2, ..., q$, compute the (Euclidian) distance between any pair of training patterns;

ii) For each class Ω_i, $i = 1, 2, ..., q$, compute the corresponding average distances and standard deviations, denoted by D_i, SD_i respectively.

iii) For each class Ω_i, $i = 1, 2, ..., q$, compute the corresponding confidence intervals $I_{\Omega_i} = (D_i - 3SD_i, D_i + 3SD_i)$ for the average distances. This intervals represent the domains of the smoothing factors σ_i.

iv) For each decision class Ω_i, $i = 1, 2, ..., q$, consider the Parzen-Cacoulos classifiers $f_i(x)$ as the corresponding parent densities. Assign (σ_i, D_i), $i = 1, 2, ..., q$.

v) In each decision class Ω_i (randomly) choose a certain vector x_i^0 and compute $f_i(x_i^0)$.

vi) (*Bayes decision rule*) Compare $f_i(x_i^0)$ and $f_j(x_i^0)$ for all $i \neq j$ following the algorithm: "IF $l_i h_i f_i(x_i^0) > l_j h_j f_j(x_i^0)$ (for all $j \neq i$) THEN $x_i^0 \in \Omega_i$ ELSE IF $l_i h_i f_i(x_i^0) \leq l_j h_j f_j(x_i^0)$ (for some $j \neq i$) THEN $x_i^0 \notin \Omega_i$".

vii) (*Measuring the classification accuracy. Updating counter*) For each (fixed) decision class Ω_i consider the 3-valued logic: TRUE -if $l_i h_i f_i > l_j h_j f_j$ (for all $j \neq i$), UNKNOWN -if $l_i h_i f_i = l_j h_j f_j$ (for some $j \neq i$) and FALSE -otherwise. Initially, each of the three variables is set to

zero. Whenever a truth value is obtained, the corresponding variable is incremented with step size 1.

viii) The cost function is given by the sum of training patterns that are classified in the right way.

ix) Repeat step 5 for another choice for x_i^0 in Ω_i until all of them are chosen. Increment counter.

x) Repeat step 5 for all vectors x_j^0 in Ω_j for all $j \neq i$. Increment counter.

xi) (*Searching for optimal smoothing parameter*) Generate in each confidence interval I_{Ω_i} a number of N random dividing points $\{P_1,\ P_2, ...,\ P_N\}$, uniformly distributed in I_{Ω_i}. Repeat step 5 by assigning $\sigma_i = P_k,\ k = 1,\ 2, ...,\ N$ for each $i = 1,\ 2, ...,\ q$.

xii) Find the maximum of the cost function.

Output. $\sigma_i,\ i = 1,\ 2, ...,\ q$, corresponding to the maximum of the cost function, represent the optimal values of the smoothing parameters $\sigma's$ for each decision category $\Omega_i,\ i = 1,\ 2, ...,\ q$.

Note. It is well-known that, on the one hand, the health care modelling domain frequently encounters situations of non-numeric data (e.g. nominal data, ordinal data, images, multimedia data, even data collected from WWW) and, on the other hand, PNNs do not tend to perform well with such a data. Moreover, in the use of complex patterns in the health care area, weights for the attributes may be incorporated, in order to highlight the importance of each attribute. Under these circumstances, the Euclidian distance used in the PNN algorithm does not work correctly any more. Fortunately, there are methods to deal with these problems [Bishop, 1995]. One of the simplest approaches consists in using a mixed-weighted measure of similarity instead of the Euclidian distance [Gorunescu, 2003]. Such a measure allows us to compute distances between training patters consisting in numerical and non-numerical attributes (e.g. images) and taking into account the significance of each attribute in the decision process.

4 PNN application to hepatic cancer diagnosis

The PNN-based decision model was applied to classify a group of individuals into a certain categories of diagnosis in the area of hepatic diseases. This application might be seen as a case-control study investigating a way of selecting people with liver cancer (HCC) -the cases, using comparable persons who do not have this disease (the controls). It has been suggested [Ibrahim and Spitzer, 1979] that a case-control study requires at least two control groups to minimize the possibility of accepting a false result; the rationale is that if the same result is not achieved in the two case-control comparisons, both the apparent results are suspect. In our application there is a case group (HCC) and three control groups (CH), (LC) and (HP). Since PNNs are particularly useful for classification problems with more than two outputs, we have enlarged the previous case-control study in order to classify

people in four diagnosis group: healthy people (HP), chronic hepatitis (CH), liver cirrhosis (LC) and hepatic cancer (HCC), instead of persons developing hepatic cancer vs. persons who do not have the disease.

The PNN-based classification algorithm has been applied to data in order to classify the initial group of individuals into four categories, depending on the diagnosis type: Ω_1 = HCC, Ω_2 = LC, Ω_3 = CH and Ω_4 = HP. Each person in the data set is represented by a 15-dimensional vector $\mathbf{x} = (x_1,\ x_2, ...,\ x_{15})$ where the components represent some of the most important characteristics leading to the right medical diagnosis. Concretely, x_1 = TB (total bilirubin), x_2 = DB (direct bilirubin), x_3 = IB (indirect bilirubin), x_4 = AP (alkaline phosphatase), x_5 = GGT (gamma glutamyl transpeptidase), x_6 = LAP (leucine amino peptidase), x_7 = AST (aspartate amino transferase), x_8 = ALT (alanine amino transferase), x_9 = LDH (lactic dehydrogenase), x_{10} = PI (prothrombin index), x_{11} = GAMMA, x_{12} = ALBUMIN, x_{13} = GLYCEMIA, x_{14} = CHOLESTEROL and x_{15} = AGE. An example of such training data vector related to hepatic cancer is the following one: (6.97, 3.04, 3.93, 438, 279, 182, 135, 52, 95, 450, 3.6, 80, 1.2, 56, 1).

The model was fitted to real data consisting of 299 individuals (both patients and healthy people) from the Department of Internal Medicine, Division of Gastroenterology, University Emergency Hospital of Craiova, Romania. This group of individuals consists of 60 patients with chronic hepatitis (CH), 179 patients with liver cirrhosis (LC), 30 patients with hepatocellular carcinoma (HCC) and 30 healthy people (HP).

5 Experimental results

It is worth to mention that we have used only raw data without any previous data checking or data preparation (some errors in recording data or the existence of certain outliers is thus possible); moreover, no data screening has been performed [Altman, 1990]. The goal of such an approach is to verify the robustness of the PNN technique to learn from raw data.

The key to obtain a good classification using PNNs is to optimally estimate the two parameters of the Bayes decision rule, the misclassification costs and the prior probabilities. Unfortunately, there is no definitive science to obtain them and must be assigned as a specific part of the problem definition. In our practical experiment we have estimate them heuristically. Thus, as concerns the costs parameters, we have considered them depending on the average distances D_i, inversely proportional, that is $l_i = 1/D_i$; in this case the accuracy rate for N = 450 was about 90%. As concerns the prior probabilities, they measure the membership probability in each group and, thus, we have considered them equal to each group size, that is $h_i = m_i$.

To avoid overfitting, the data set was randomly partitioned into two sets: the training set and the validation set. A number of 254 persons (85%) of the initial group was withheld from the initial group for the smoothing factor

adjustment (the training process). Once optimal smoothing parameters $\sigma's$ for each decision category were obtained using the training set, the trained PNN was applied to the validation set (the remaining 45 persons). Since we have used raw data to perform the PNN algorithm and to avoid the criticism of some people against the Monte Carlo method due to the fact the smallness of the error of method is only ensured with a certain probability, we have repeated 10 times the above procedure to diminish the outliers influence and a possible Monte Carlo technique weakness.

We have use the Java package for the algorithm implementation. What is important about the Java implementation of the program is that all data about patients collected by physicians can, at any time, be added, modified or deleted, with no change in the source of the program whatsoever. That is so because for the processing of the data we have used JDBC (Java Database Connectivity). Thus the program is connected to a database and the records of the specific table of this database can always be updated by the users themselves (in MS Access or MS Excel) with no further worries concerning the applicability of the program.

The experimental results are shown in Table 1 and Table 2. Table 1 presents the accuracy rates for both the training process and for the validation process.

Training accuracy rate (%)	Validation accuracy rate (%)
97.32	92.22
85.28	88.88
85.61	95.55
95.65	93.88
91.60	92.00
89.62	93.66
89.28	93.77
90.26	92.22
87.94	91.22
88.29	93.33
Average accuracy 90.10	92.67

Table 1. PNN classifier: experimental results

When the PNN was applied to the training process, the sensitivity analysis indicated that the proportion of the patients correctly diagnosed was (average) 90.10%.

When the PNN was applied to the validation data set, which was not subjected to neural network training, the proportion was (average) 92.67%.

The general predictive abilities of the PNN with the validation data set is particularly positive, given the fact that the validation data were not used in the training of the neural network.

In Table 2 we have considered the 3-valued logic: TRUE, FALSE and UNKNOWN and we have displayed the accuracy rates obtained during the validation process related to this classification, that is the percentage of patients correctly classified, incorrectly classified and unclassified.

Correctly classified patients	Incorrectly classified patients	Unclassified patients
92.22	6.78	1.00
88.88	8.12	3.00
95.55	4.45	0.00
93.88	6.12	0.00
92.00	7.00	1.00
93.66	6.34	0.00
93.77	6.23	0.00
92.22	7.78	0.00
91.22	8.78	0.00
90.10	7.90	2.00

Table 2. PNN classifier: classification correctness

We see that the unclassified cases represent at most 3% of the whole number of patients and in 60% of the computer program running we obtained no unclassified cases.

6 Conclusion and further work

In this paper we have developed and demonstrated the applicability and suitability of a PNN-based model for decision-making in the hepatic diagnosis process. PNNs learn by examples so the details of how to recognize the disease are not needed. What is needed is a set of examples that are representative of all the variations of the disease. We used raw data (the only data available for the experiment) and we obtained reliable results proving the PNNs ability and flexibility to learn by raw examples.

A problem to deal with in PNNs applications is the data set size. The number of cases required for PNN training frequently presents difficulties. As the number of variables increases, the number of cases required increases nonlinearly, so that with a fairly small number of variables a huge number of cases are required. In our experiment we used 299 cases with 15 variables. Further works should perform a heuristic study relating the number of variables to the number of cases.

In comparison with other PNN approaches related to the diagnosis process, the accuracy of this technique is competitive. For instance, in predicting ascites in broilers based on minimally invasive inputs [Roush *et al.*, 1997], a validation rate accuracy of 95% was reported. At the same time, a validation

accuracy rate of 92.3% was reported in estimating the mortality risk following cardiac surgery [Orr, 1997].

Although the early diagnosis of liver cancer in liver cirrhosis is based on biochemical tests, modern approaches also use imaging tests (i.e. transabdominal ultrasound and/or spiral computed tomography). Therefore, another way to enlarge this heuristic approach in medical research is represented by the replacement of the Euclidian distance with a general mixed-weighted measure of similarity. Such an approach will strengthen the decision process by using much more attributes of the training patterns.

Clearly, much work still needs to be done to improve this methodology and to apply it to other health care classification problems.

References

[Altman, 1990]D.G. Altman. *Practical statistics for medical research*. Chapman and Hall, 1990.

[Bishop, 1995]C. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[Cacoulos, 1966]R. Cacoulos. Estimation of a multivariate density. *Ann. Inst. Stat. Math.(Tokyo)*, 18, pages 179–189, 1966.

[Gorunescu, 2003]F. Gorunescu. Measuring similarities: an application to the chromosomes supervised selection. *Research Notes in Artificial Intelligence and Data Communications*, 103, pages 56–66, 2003.

[Ibrahim and Spitzer, 1979]M.A. Ibrahim and W. Spitzer. The case-control study: the problem and the prospect. *J. chron. Dis.*, 32, pages 130–144, 1979.

[Lin *et al.*, 2002]F. Lin, C. Chiu and S. Wu. Using Bayesian Networks for Discovering Temporal-State Transition patterns in Hemodialysis. *Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS-35.02), 0-7695-1435-9/02 © 2002 IEEE*, 2002.

[Norton *et al.*, 2001]I.D. Norton, Y. Zheng, M.S. Wiersema, J. Greenleaf, J. Clain and E. Dimagno. Neural network analysis of EUS images to differentiate between pancreatic malignancy and pancreatitis. *Gastrointest Endosc.*, 54(5), pages 625–629, 2001.

[Orr, 1997]R.K. Orr. Use of a Probabilistic Neural Network to Estimate the Risk of Mortality after Cardiac Surgery. *Med. Decis. Making*, 17, pages 178–185, 1997.

[Parzen, 1962]E. Parzen. On estimation of a probability density function and mode. *Ann. Math. Stat.*, 33, pages 1065–1076, 1962.

[Roush *et al.*, 1997]W.B. Roush, T.L. Cravener, Y.K. Kirby and R.F. Wideman. Probabilistic Neural Network Prediction of Ascites in Broilers Based on Minimally Invasive Physiological Factors *Poultry Science*, 76, pages 1513–1516, 1997.

[Specht, 1988]D.F. Specht. Probabilistic neural networks for classification mapping or associative memory. *Proceedings IEEE International Conference on Neural Networks*, 1, pages 525–532, 1988.

[Specht, 1990]D.F. Specht. Probabilistic neural networks. *Neural Networks*, 3, pages 109–118, 1990.

[Taktak *et al.*, 2004]A. Taktak, A. Fisher and B. Damato. Modelling survival after treatment of intraocular melanoma using artificial neural networks and Bayes theorem. *Phys. Med. Biol.*, 49, pages 87–98, 2004.