INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS

VOL. I

IJCNN-91-SEATTLE

International Joint Conference on Neural Networks

July 8-12, 1991
Washington State
Convention & Trade Center
Seattle, WA

Volume I

Co-sponsored by the
International Neural Network Society
and the
Institute of Electrical and Electronics Engineers, Inc.

TABLE OF CONTENTS

VOLUME I

APPLICATIONS I *Oral Presentations*

APPLICATIONS II *Oral Presentations*

APPLICATIONS III *Oral Presentations*

TABLE OF CONTENTS CONTINUED

VOLUME I

CONTINUED TABLE OF CONTENTS

VOLUME I

TABLE OF CONTENTS CONTINUED

VOLUME I

APPLICATIONS II *Poster Presentations*

CONTINUED TABLE OF CONTENTS

VOLUME I

TABLE OF CONTENTS CONTINUED

VOLUME I

ELECTRONIC & OPTICAL IMPLEMENTATIONS *Posters*

IMAGE PROCESSING & VISION I *Oral Presentations*

VOLUME I

TABLE OF CONTENTS CONTINUED

VOLUME I

LEARNING & GENERALIZATION I *Oral Presentations*

CONTINUED TABLE OF CONTENTS

VOLUME I

LEARNING & GENERALIZATION II *Oral Presentations*

LEARNING & GENERALIZATION III *Oral Presentations*

The Early Diagnosis of Heart Attacks: A Neurocomputational Approach

Robert F Harrison, Stephen J Marshall and R Lee Kennedy
Department of Control Engineering
University of Sheffield
Sheffield S1 3JD
England

Abstract

A multi-layered Perceptron is trained to diagnose the presence, or otherwise, of acute myocardial infarction (heart attack) in patients admitted to an emergency unit with acute chest pain. Two learning algorithms, based on mean-square-error and the log-likelihood function, are compared. Their performance does not differ significantly but the latter rule converges much more rapidly.

Performance in excess of that of the admitting clinicians is achieved for a number of performance indicators, and a protocol for combining the network's diagnosis with that of the clinician is proposed. This results in further improvements in performance, indicating that the MLP can act as a useful decision aid in an emergency context.

Introduction

The motivation behind this work is to demonstrate the effectiveness of neural networks in clinical decision support. To this end we have chosen a very widespread problem which demands urgent diagnosis. Suspected heart attack is the most common cause of admissions to Accident and Emergency Departments in the UK and the cause of over 1.5 million admissions annually, in the US [1]) In order that the resulting network should be a useful clinical aid we have restricted the input data to only those items which are collected in a normal case history.

The diagnosis of Acute Myocardial Infarction (AMI), or heart attack, is a complex decision making process. Patients are admitted to a hospital's emergency department suffering from chest pain, and the clinician must form a diagnosis accurately and quickly if the patient is to receive the correct treatment. Typically 20-30 % of emergency hospital admissions are of patients with possible AMI and fewer than half of those admitted to a Coronary Care Unit prove to have AMI.

Difficulties in the accurate diagnosis of AMI arise for many reasons. Patient symptoms are often atypical and can have many different causes, *eg* strained chest muscles or stomach ulcers. There is no quick reliable test to aid diagnosis, the electrocardiograph (ECG) is the only diagnostic tool available at presentation, and this often fails to show diagnostic changes. There is, however, a standard test that does give the correct diagnosis, but this involves the measurement of enzyme and ECG changes over a period of 24 to 48 hours. This gives a reference against which the clinician's performance can be measured, but is of no use in the early stages of diagnosis.

In addition to deciding whether to admit the patient to hospital for observation, or to a Coronary Care Unit (CCU) a third decision of whether or not to administer thrombolytic therapy can be made. To be of use this must be administered within about 6 hours of the heart attack. It benefits patients by reducing mortality and heart tissue damage [2], but can be very expensive and there are risks associated with inappropriate use [3]. In the US it has recently been estimated that of those requiring thrombolytic therapy only 10% receive it [4]

There is an obvious need therefore for a clinical decision aid to assist the clinician in making a fast, accurate diagnosis. A neurocomputational approach is well suited to this problem. First, the frequent occurrence of AMI enables a large amount of patient data to be collected relatively easily. Second, the standard test gives a 'correct' diagnosis for each patient, this can be used as a target for training a network such as the Multi-Layer Perceptron (MLP) used here. Third, the difficulty of extracting explicit rules for the diagnosis makes the application of more traditional expert system techniques difficult and time consuming [5] [6]. Such methods also suffer from portability problems in that refining a system to work well in a different geographical location where, for instance referral procedures make be markedly different, may take a long

0-7803-0164-1/91/0000-0001$01.00©1991 IEEE

time. Conventional statistical techniques have also been applied to this problem but none has found wide acceptance. To the authors' knowledge, only one other study of this problem has been conducted using neural networks [7] but it achieved an unacceptably low level of accuracy. We are unable to ascertain a reason for this

In our work we interpret the network inputs as probabilities of the presence of a particular symptom and its output as the probability that AMI is present. Baum and Wilczek [8] propose that by treating the external variables in such a way and varying the weights in the gradient direction of log likelihood during training, the probabilistic interpretation is allowed. In addition learning is made more efficient. To determine any advantages of this method experiments were performed comparing the performance of networks trained using the least-mean-square backpropagation algorithm and the maximum likelihood modification.

In order to provide the clinician with information about how a particular diagnosis was arrived at, we have investigated the sensitivity of the network output to individual symptoms. We believe that such "contribution analyses" will be invaluable in overcoming objections to the use of neural networks on the grounds that they are "black-boxes" .

The log-likelihood cost function

It is claimed by Baum and Wilczek [8] that backpropagation can be put on a satisfactory conceptual footing and probably be made more efficient by defining the values of output and input neurons as probabilities, and replacing the mean-square-error (MSE) by the log-likelihood function (LL) resulting in a maximum likelihood (ML) rather than a least mean-square (LMS) optimization. From our point of view, treatment of symptoms and diagnoses as probability of occurrence is most appealing and has led us to compare the two learning algorithms.

Given a set of N input vectors, s^{μ} ; $\mu \in [0, N]$, and their associated target values t^{μ}, backpropagation attempts to adjust the network weights so as to minimize the MSE, E_{MS}, between the output and target values as defined by $E_{MS} = \sum_{\mu,j}(t_j^{\mu} - o_j^{\mu})^2$ where o_j^{μ} is the output of the j-th node when s^{μ} is presented as input to the network. This formulation for MSE can be replaced with the LL function, E_{LL} given by $E_{LL} = -\sum_{\mu,j}\left\{t_j^{\mu}\log o_j^{\mu} + (1-t_j^{\mu})\log(1-o_j^{\mu})\right\}$. By varying the weights in the gradient direction of this function (found by the chain-rule of differentiation) the backpropagation algorithm generalises immediately from minimizing MSE to maximizing LL and hence likelihood. Holt and Semani [9] have also investigated this and conclude that the use of the LL formulation can significantly reduce training times in the network, and can increase the dynamic ranges of the initial weights over which proper convergence can be achieved. They also note that despite these apparent advantages the LL approach has not found wide acceptance in applications. Here we investigate any practical differences between the two methods when applied to the AMI problem.

The network structure

The network architecture used here is the multi-layered Perceptron (MLP). One intermediate layer of neurons is used which is fully interconnected with the 53 neurons of the input layer and the single neuron of the output layer. The optimum size of the hidden layer is notoriously difficult to define and here guidance is obtained using the algorithm proposed in [10], with a figure of 18 being chosen.

Contribution analysis

One of the major criticisms of the MLP is its inability to explain the decision it has reached. The large number of connections and the nonlinearities in the hidden layers obscures the rôle of any particular input in the final diagnosis. In an attempt to overcome this problem we have devised a method to analyse the effect of each input factor on the output. This is done simply by calculating the sensitivity of the output to any individual input *ie* the effect of a unit change in that input. In fact we calculate the sensitivity of the *net input* to each output unit to preserve dynamic-range. This is justified by the monotonicity of the output units' sigmoid functions, $\sigma(.)$. This is easily done using the chain-rule of differentiation yielding

$$S_{i,m}^{\mu} = \frac{\partial net_i^{\mu}}{\partial s_m^{\mu}} = \sum_{j\in \mathrm{H}} w_{i,j}\sigma'\left(\sum_{k\in \mathrm{I}} w_{j,k}^{\mu}s_k^{\mu}\right) w_{j,m}$$

where H and I denote the units of the hidden and input layers respectively.

The sensitivity of the i^{th} output unit to the m^{th} input unit is a function of all the inputs as we would expect. This makes for difficulties in deriving a global measure of sensitivity. However, because the sign of σ' is always positive, the sign of $S^{\mu}_{i,m}$ remains unchanged for any set of symptoms. This means that we can average over all cases. This gives a number for each input unit relating to its contribution for or against the diagnosis, serving both to highlight the important factors used to make the decision and allowing clinicians to check any suspect factors in the input to the network.

The data

The data for the study were obtained from 300 consecutive emergency referrals, with a complaint of chest pain, to the Medical Department of the Northern General Hospital, Sheffield, England. Demographic, clinical and electrocardiographic data was recorded from all patients, as was the admitting clinician's estimate of the likelihood of the patient having suffered AMI (as a percentage). Myocardial infarction was diagnosed according to the standard criterion thus supplying the "teacher" for the network. From each patient record 38 symptoms were coded as a 53 dimensional, binary vector and the target vector (diagnosis) was coded as 1 for AMI, 0 otherwise. Continuous valued variables such as age and duration of pain were coded as binary vectors eliminating logically redundant elements as described by Widrow *et al* [11]. The first 150 symptom/diagnosis pairs were used to train the network while the second 150 were used to test it.

Results

Speed of learning

To confirm the predicted increase in learning speed both the LMS and ML algorithms were trained for 1000 epochs. The learning rate of the ML network was set at four times smaller than that of the LMS net (0.01). This takes account of the fact that changes in gradient in the ML net can be more than four times greater [8] than those taking place in the LMS variant so we might expect the two systems to learn at similar rates. Figure 1 clearly demonstrates that the MSE for both nets falls off very much more quickly for the ML net. Similar, although mirror-image, curves (not shown) can be seen if LL is calculated for each network. This obviously has implications for the required computational power.

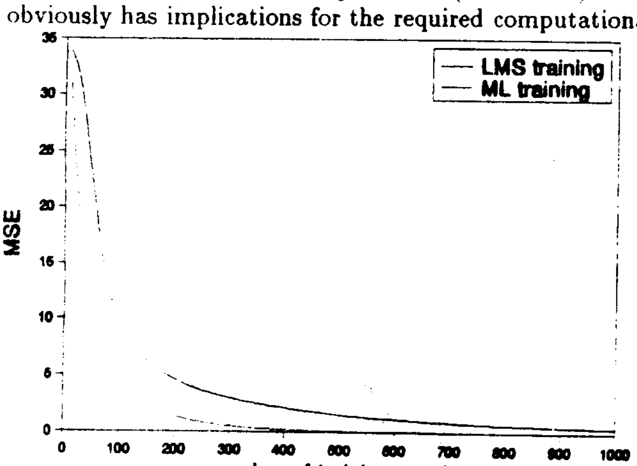

Fig. 1: MSE v no. epochs for LMS and ML training

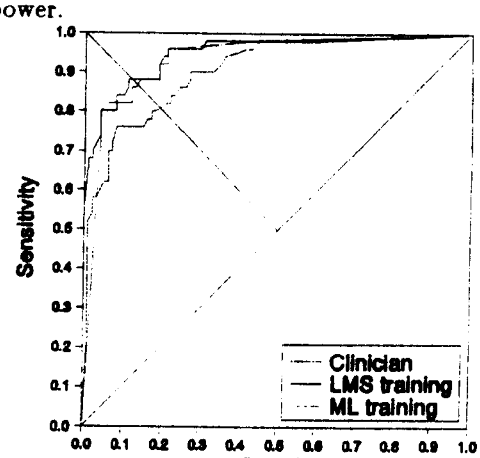

Fig. 2: ROC curve for LMS & ML nets & clinician

Diagnostic capability

The diagnostic power of the trained MLP was assessed by calculating the following performance indicators for the test set of (previously unseen) data. They are defined as follows [12]. Diagnostic accuracy: ratio of number of correct diagnoses to total number of cases; Sensitivity: ratio of number of correct ***positive*** diagnoses to the total number of patients ***with*** the disease; Specificity: ratio of number of correct ***negative*** diagnoses to the total number of patients ***without*** the disease.

In the language of decision theory sensitivity is the probability of detection and specificity, 1 - probability of a false alarm. It is easy, therefore, to plot the Receiver Operating Characteristic (ROC) curve for each of the experiments [13]. This is shown in figure 2 and includes the ROC curve for the admitting clinicians, for comparison. The intersection of the ROC curve with the semi-diagonal indicates the optimum threshold level, T_o, (*ie* sensitivity = specificity = accuracy) that each system can operate at for a simple yes/no diagnosis. Thresholds and performance are tabulated in table 1. It is clear from figure 2 that there is no significant difference between the ML and LMS algorithms and that the clinicians performed worse than either of the networks.

The objective of this work is not to develop an alternative to the clinician, but rather an aid which can make optimal use of all available data at presentation *including* the clinician's opinion. For instance, we might make the combined network and clinician system more *specific* by insisting that both diagnoses should be positive if AMI is to be diagnosed (logical AND). Such a protocol might be valid when trying to decide whether or not to administer a thrombolytic agent. A less critical decision, that of whether or not to admit the patient to CCU might benefit from the combination of network and clinician using logical OR. This says if either diagnosis is positive then AMI is present and increases *sensitivity*. A third option arises which admits a "grey" area of opinion, representing probabilities of AMI lying between an upper and a lower limit, say, 10% and 90%. Table 2 indicates a possible scoring system for combining diagnoses. Using this weighted combination of clinical opinion and output from the neural network, MI is diagnosed if a combined score of 3 or more is obtained. Notice how responsibility ultimately devolves to the clinician. It is not suggested that this protocol is in any way optimal. It was chosen arbitrarily to demonstrate how a workable system might be achieved. Clearly, a good deal of effort is required to define the way in which machine and clinician should best co-operate.

	T_o	acc	sens	spec
LMS net	0.57	0.88	0.88	0.88
ML net	0.74	0.87	0.86	0.87
clinician	0.47	0.81	0.80	0.82
LMS + clin	—	0.89	0.96	0.86
ML + clin	—	0.87	0.94	0.84

Table 1: Performance

Clinician		Network	
Prob MI (%)	Score	Prob MI (%)	Score
0—10	0	0—10	0
11—50	1	11—50	1
51—90	2	51—90	2
91—100	3	91—100	2

Table 2: Scoring system

Figure 3 shows the average contribution over all cases for each input for both the LMS and ML networks. Positive contributions add to the overall output while negative ones subtract. The six most important positive and four most important negative contributors are listed in table 3. All agree with clinical opinion, for example, the presence or absence of ECG changes are some of the major factors used by the clinician. Unfortunately space does not allow us to discuss them in detail here. The order of importance for both networks is similar although the absolute values of $S_{i,m}$ for the ML network tend to be lower. This can be explained by the fact that we have observed that the final weights in the ML net tend to be smaller than those in the LMS net.

m	$S_{i,m}$		Symptom
	ML	LMS	
48	4.9	8.2	New ST segment elevation
50	2.8	4.1	ST or T wave suggest isch.
47	2.1	3.4	Patient hypoperfused
49	1.9	2.7	New pathological Q waves
13	1.7	2.4	Other risk factor
26	1.7	2.9	Associated with Sweating
31	-2.2	-3.2	Episodic pain
42	-1.9	-3.1	History of Angina
10	-1.6	-2.4	Family history of isch. HD
21	-1.2	-1.8	Pain worse on inspiration

Table 3: Most important contributors

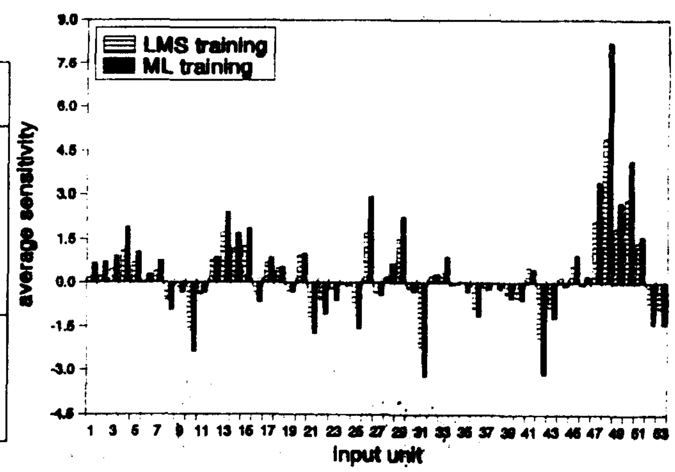

Fig. 3: Average sensitivity of output on inputs

Conclusions

For the present study we conclude that the performance of the ML and LMS trained networks does not differ significantly. However, there is a distinct advantage in terms of speed of learning when the ML algorithm is used.

For a computer-based diagnostic aid to be of value it must be statistically valid, diagnostically accurate and its use must enhance the clinician's diagnostic performance. In addition it must be readily usable in the clinical setting. In this study we have not rigourously addressed the question of statistical validity but we have developed a system whose performance can exceed that of experienced physicians. Our system is readily usable on a portable computer, could be developed as a hand held instrument and gives an instant prediction of the likelihood that a patient has sustained AMI.

We have also shown how our system could be used to enhance the clinician's judgement and to indicate the importance of those factors which influence the decision making process. The preliminary results in this latter area are encouraging and we believe that this will be a fruitful line of enquiry. As with all clinical decision aids though, the true performance of the system can only be assessed by a formal, clinical trial.

References

[1] Pozen, M *et al* (1984) A predictive instrument to improve coronary care unit admission practices in acute ischaemic heart disease: a prospective multi-centre clinical trial *New England J Med* **310** 1273-1278.

[2] ISIS-2 (Second International Study of Infarct Survival) Collaborative Group (1988) Randomized trial of intravenous streptokinase, oral aspirin, both or neither, among 17,187 cases of suspected acute myocardial infarction *Lancet* **ii** 349-360.

[3] Rawles, J and Julian, D (1989) Role of the general practitioner in managing patients with myocardial infarction impact of thrombolytic treatment *Brit Med J* **299** 555-557.

[4] Kennedy, J W. (1990) Expanding the use of thrombolytic therapy for AMI *Annals of Internal Med* **113** 907-908.

[5] Fattu, J, Blomberg, D and Patrick, E (1987) CONSULT learning system applied to early diagnosis of chest pain *Proc* 11th *Annual Symposium on Computer Applications in Medical Care* 71-77.

[6] Emerson, P *et al* (1988) The development of ACORN, an expert system enabling nurses to make admission decisions about patients with chest pain in an accident and emergency department *Medical Informatics in Clinical Medicine, Nottingham* 37-40.

[7] Hart, A and Wyatt, J (1989) Connectionist models in medicine: an investigation of their potential *Proc European Conf on AI in medicine, London* 115-124.

[8] Baum, E B and Wilczek F (1988) Supervised learning of probability distributions by neural networks *IEEE Conf. on Neural Information Processing Systems, ed. Anderson, New York: American Inst. Physics* 52-61.

[9] Holt, M J J and Semnani, S (1990) Convergence of back-propagation in neural networks using a log-likelihood cost function *Electronic Letters* **26** 1964-1965.

[10] Gutierrez, M, Wang, J and Lipschik, G (1989) Estimating hidden units for two-layer Perceptrons *Proc IEE 1st Int Conf on Artificial Neural Networks* 120-124.

[11] Widrow, B, Gupta, N and Maitra, S (1973) Punish/reward: learning with a critic in adaptive threshold systems *IEEE Trans on Systems, Man and Cybernetics* **SMC 3** 455-465.

[12] Harrison, R F, Kennedy, R L, Marshall, S J (1991) The multi-layer perceptron as an aid to the early diagnosis of myocardial infarction *Proc Third European Conference on Artificial Intelligence in Medicine* (To appear).

[13] Meistrell, M L (1990) Evaluation of neural network performance by receiver operating characteristic (ROC) analysis: examples from the biotechnology domain *Computer Methods and Programs in Biomedicine* **32** 73-80.