Module 3

Multilayer Neural Networks and the Backpropagation Algorithm

Prof. Marzuki Bin Khalid
CAIRO
Fakulti Kejuruteraan Elektrik
Universiti Teknologi Malaysia
marzuki@utmkl.utm.my

Module 3 Objectives

- To understand what are multilayer neural networks.
- To understand the role and action of the logistic activation function which is used as a basis for many neurons, especially in the backpropagation algorithm.
- To study and derive the backpropagation algorithm.
- To learn how the backpropagation algorithm is used to solve a simple XOR problem and character recognition application.
- To try some hands-on exercises for understanding the backpropagation algorithm.

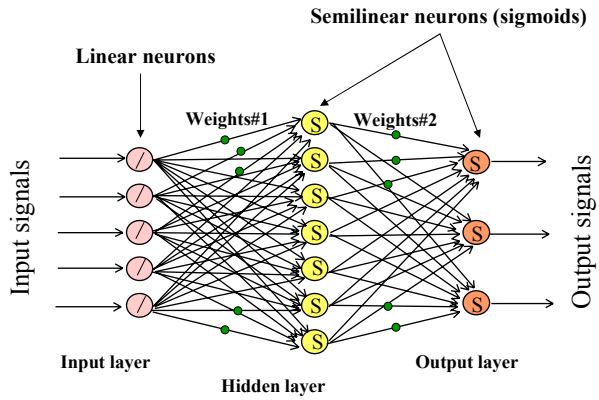
Module Contents

3.0 Multilayer Neural Networks and The Backpropagation (BP) Algorithm

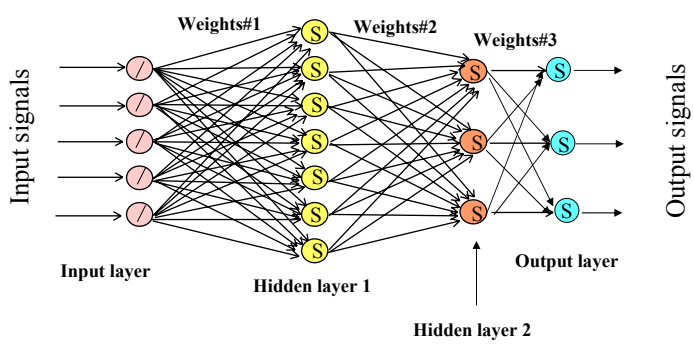
3.1 Multilayer Neural Networks

- Multilayer neural networks are feedforward ANN models which are also referred to as multilayer perceptrons.
- The addition of a hidden layer of neurons in the perceptron allows the solution of nonlinear problems such as the XOR, and many practical applications (using the backpropagation algorithm).
- However, the difficulty of adaptation of the weights between the hidden and input layers of the multilayer perceptrons have dampen such architecture during the sixties.
- With the discovery of the **backpropagation algorithm** by Rumelhart, Hinton and Williams in 1985, the adaptation of the weights in the lower layers of multilayer neural networks are now possible.

- The researchers proposed the use of semilinear neurons with differentiable activation functions in the hidden neurons referred to as logistic activation functions (or sigmoids) which allows the possibility of such adaptation.
- A multilayer neural network with one layer of hidden neurons is shown below (also called a two-layer network).



- An example of a three-layered multilayer neural network with two-layer of hidden neurons is given below.



3.2 The Logistic Activation (Sigmoid) Function

- Activation functions play an important role in many ANNs.
- In the early years, their role is mainly to fire or unfire the neuron.
- In new neural network paradigms, the activation functions are more sophisticatedly used.
- Many activation functions used in ANNs nowadays produce a continuous value rather than discrete.
- One of the most popular activation functions used is the logistic activation function or more popularly referred to as the sigmoid function.
- This function is semilinear in characteristic, differentiable and produces a value between 0 and 1.

- The mathematical expression of this sigmoid function is:

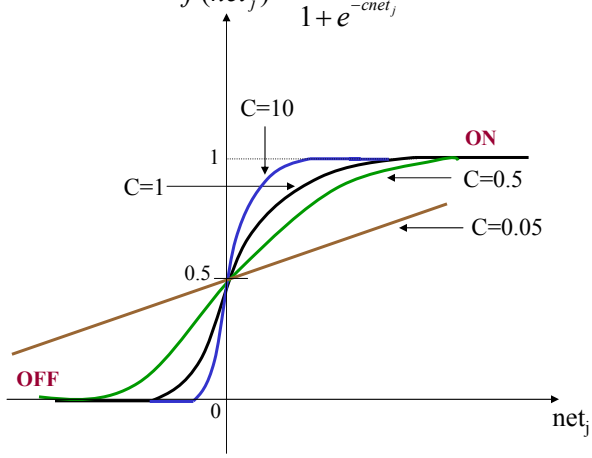
$$f(net_j) = \frac{1}{1 + e^{-cnet_j}}$$

where c controls the firing angle of the sigmoid.

- When c is large, the sigmoid becomes like a threshold function and when is c small, the sigmoid becomes more like a straight line (linear).
- When c is large learning is much faster but a lot of information is lost, however when c is small, learning is very slow but information is retained.
- Because this function is differentiable, it enables the B.P. algorithm to adapt the lower layers of weights in a multilayer neural network.

- The sigmoid activation function with different values of c.

$$f(net_j) = \frac{1}{1+e^{-cnet_j}}$$



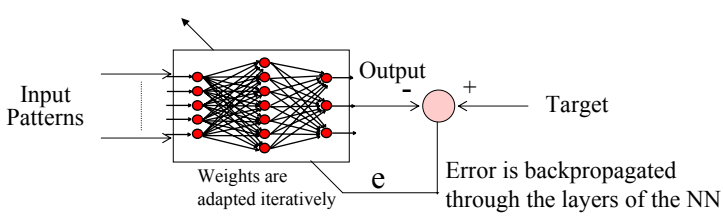
3.3 The Backpropagation (BP) Algorithm

- The BP algorithm is perhaps the most popular and widely used neural paradigm.
- The BP algorithm is based on the generalized delta rule proposed by the PDP research group in 1985 headed by Dave Rumelhart based at Stanford University, California, U.S.A..
- The BP algorithm overcame the limitations of the perceptron algorithm.
- Among the first applications of the BP algorithm is speech synthesis called NETalk developed by Terence Sejnowski.
- The BP algorithm created a sensation when a large number of researchers used it for many applications and reported its successful results in many technical conferences.

3.3.1 Learning Mode

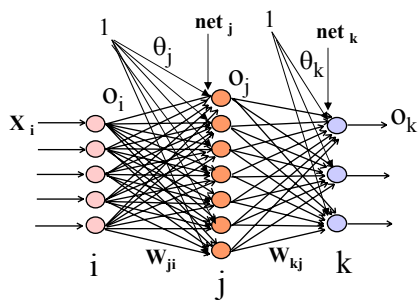
- Before the BP can be used, it requires target patterns or signals as it a supervised learning algorithm.
- Training patterns are obtained from the samples of the types of inputs to be given to the multilayer neural network and their answers are identified by the researcher.
- Examples of training patterns are samples of handwritten characters, process data, etc. following the tasks to be solved.
- The configuration for training a neural network using the BP algorithm is shown in the figure below in which the training is done offline.
- The objective is to minimize the error between the target and actual output and to find Δw.

- The error is calculated at every iteration and is backpropagated through the layers of the ANN to adapt the weights.
- The weights are adapted such that the error is minimized.
- Once the error has reached a justified minimum value, the training is stopped, and the neural network is reconfigured in the recall mode to solve the task.



3.3.2 The Generalized Delta Rule (G.D.R.)

- The objective of the BP algorithm, like in other learning algorithms, is to find the next value of the adaptation weights (Δw) which is also known as the G.D.R..
- We consider the following ANN model:



- What we need is to obtain the following algorithm to adapt the weights between the output (k) and hidden (j) layers:

$$\boxed{\Delta W_{kj}(t+1) = \eta \delta_k O_j + \alpha \Delta W_{kj}(t)} \longleftarrow \text{G.D.R.}$$

where the weights are adapted as follows:

$$W_{kj}(t+1) = W_{kj}(t) + \Delta W_{kj}(t+1)$$

- where t is the iteration number and δ_k is the error signal between the output and hidden layers:

$$\delta_k = O_k(1 - O_k)(t_k - O_k)$$

- Adaptation between input (i) and hidden (j) layers :

$$\boxed{\Delta W_{ji}(t+1) = \eta\delta_j O_i + \alpha\Delta W_{ji}(t)} \longleftarrow \text{G.D.R.}$$

The new weight is thus:

$$W_{ji}(t+1) = W_{ji}(t) + \Delta W_{ji}(t+1)$$

and the error signal through layer j is:

$$\delta_j = O_j(1 - O_j)\sum_k \delta_k W_{kj}$$

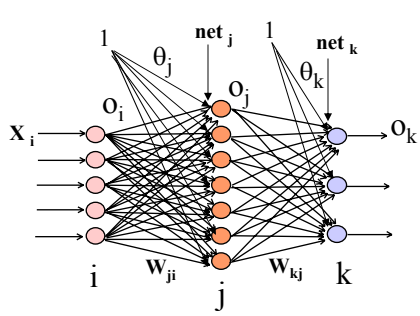
Note that

$$\boxed{\begin{aligned} O_j &= f(net_j) = \frac{1}{1+e^{-net_j}} \\ net_j &= \sum_j W_{ji} O_i + \theta_j \end{aligned}}$$

3.3.3 Derivation of the BP Algorithm

- The BP algorithm involves two phases:
  (1). Forward propagation
  (2). Backward propagation

- Before we derive the BP algorithm, we first need to consider the ANN model to be used in the derivation which is given as shown below.



- Our ANN model has the following assumptions:
  - A two-layer multilayer NN model, i.e. with 1 set of hidden neurons.
  - Neurons in layer i are fully connected to layer j and neurons in layer j are fully connected to layer k.
  - Input layer neurons have linear activation functions and hidden and output layer neurons have logistic activation functions (sigmoids).
  - Assume the firing angle of the logistic activation function, c=1.
  - Bias weights are used with bias signals of 1 for hidden (j) and output layer (k) neurons.
  - In many ANN models, bias weights (θ) with bias signals of 1 are used to speed up the convergence process.

  - The learning parameter is given by the symbol η and is usually fixed a value between 0 and 1, however, in many applications nowadays an adaptive η is used.
  - Usually η is set large in the initial stage of learning and reduced to a small value at the final stage of learning.
  - A momentum term α is also used in the G.D.R. to avoid local minimas.
- The ANN model can be looked upon as a distributed parameter system, and in the derivation of the BP algorithm, partial derivatives and chain rules are often used such as that given below.

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial u} \cdot \frac{\partial u}{\partial x}$$

3.3.3.1 The Forward Propagation step

- The output of neuron j can be expressed as :

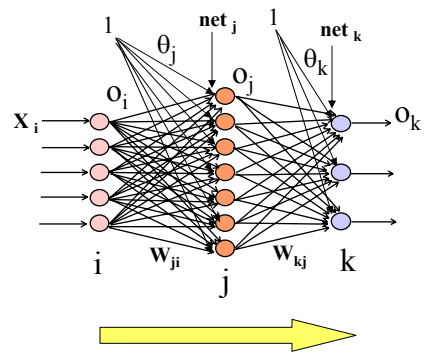
$$O_j = f(net_j) = \frac{1}{1+e^{-net_j}}$$

where $$net_j = \sum_i W_{ji} O_i + \theta_j$$

- and the output of neuron k which is the ANN output is :

$$O_k = f(net_k) = \frac{1}{1+e^{-net_k}}$$

where $$net_k = \sum_j W_{kj} O_j + \theta_k$$



3.3.3.2 Adaptation of the Weights between output layer k and hidden layer j

- Adaptation of weights between the output and hidden layers starts from

$$\Delta W_{kj} \propto -\frac{\partial E}{\partial W_{kj}} = -\eta \frac{\partial E}{\partial W_{kj}}$$

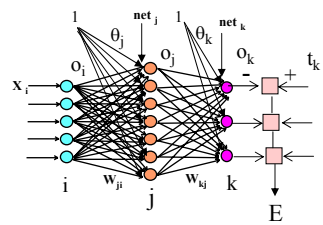
- The negative sign shows a gradient descent approach where the next error is always less.

$$\frac{\partial E}{\partial W_{kj}} = \underbrace{\frac{\partial E}{\partial net_k}}_{\delta_k} \frac{\partial net_k}{\partial W_{kj}}$$

δ_k .. which is the error signal between layer k and j

$$\delta_k = \frac{\partial E}{\partial net_k} = \frac{\partial E}{\partial O_k} \frac{\partial O_k}{\partial net_k}$$

$$\frac{\partial E}{\partial W_{kj}} = \frac{\partial E}{\partial O_k} \frac{\partial O_k}{\partial net_k} \frac{\partial net_k}{\partial W_{kj}}$$



$$\frac{\partial E}{\partial W_{kj}} = \frac{\partial E}{\partial O_k} \frac{\partial O_k}{\partial net_k} \frac{\partial net_k}{\partial W_{kj}}$$

$$\frac{\partial E}{\partial O_k} = ?$$

$$\frac{\partial O_k}{\partial net_k} = ?$$

$$\frac{\partial net_k}{\partial W_{kj}} = ?$$

- The adaptation of the weights between layer k and j is:

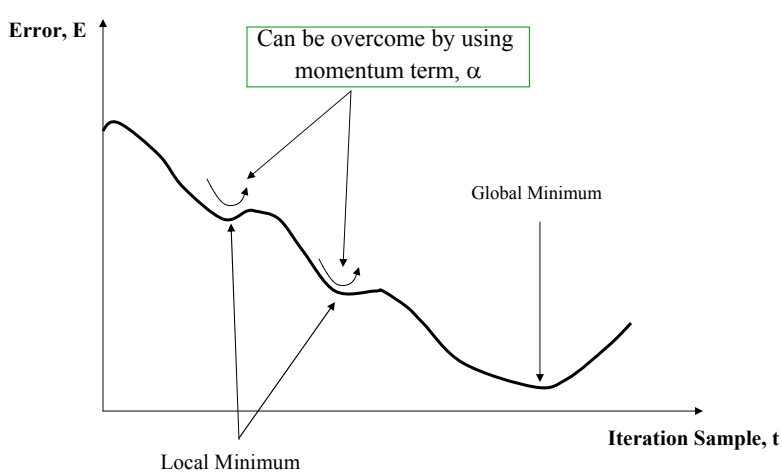
$$\Delta W_{kj} = \eta \underbrace{(t_k - O_k) O_k (1 - O_k)}_{\delta_k} O_j$$

- For better convergence a momentum term is added

$$\Delta W_{kj}(t) = \eta \delta_k O_j + \alpha \Delta W_{kj}(t-1)$$

… where t is the iteration number.

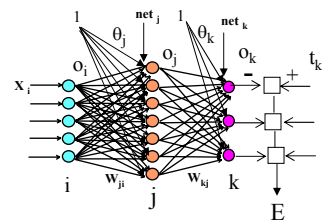
The Gradient Descent Approach is trying to find the global minimum



3.3.3.3 <u>Adaptation of weights between hidden (j) and input (i) layers</u>

Similarly, $\Delta W_{ji} \propto -\dfrac{\partial E}{\partial W_{ji}} = -\eta \dfrac{\partial E}{\partial W_{ji}}$

By chain rule, $\dfrac{\partial E}{\partial W_{ji}} = \dfrac{\partial E}{\partial net_j} \dfrac{\partial net_j}{\partial W_{ji}}$



Further expand, $\dfrac{\partial E}{\partial W_{ji}} = \dfrac{\partial E}{\partial O_j} \dfrac{\partial O_j}{\partial net_j} \dfrac{\partial net_j}{\partial W_{ji}}$

$$\frac{\partial E}{\partial W_{ji}} = \frac{\partial E}{\partial O_j} \frac{\partial O_j}{\partial net_j} \frac{\partial net_j}{\partial W_{ji}}$$

$$\frac{\partial E}{\partial O_j} = ?$$

$$\frac{\partial O_j}{\partial net_j} = ?$$

$$\frac{\partial net_j}{\partial W_{ji}} = ?$$

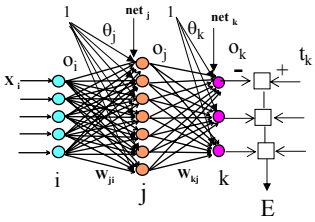
To solve $\dfrac{\partial E}{\partial O_j}$

Expand by chain rule, we obtain: $\dfrac{\partial E}{\partial O_j} = \dfrac{\partial E}{\partial net_k}\dfrac{\partial net_k}{\partial O_j}$

$\dfrac{\partial E}{\partial net_k} = \delta_k$ … has already been solved

Since, $net_k = \sum_{j=1}^{j=J} W_{kj} O_j + \theta_j$

Thus, $\dfrac{\partial net_k}{\partial O_j} = W_{kj}$



Hence $$\frac{\partial E}{\partial W_{ji}} = \underbrace{\sum_{k=1}^{k=K} \delta_k W_{kj} O_j \left(1 - O_j\right)}_{\delta_j} O_i$$

Thus, adaptation is

$$\Delta W_{ji}(t) = \eta \delta_j O_i + \alpha \Delta W_{ji}(t-1)$$

where $\delta_j = \dfrac{\partial E}{\partial net_j} = \sum_{k=1}^{k=K} \delta_k W_{kj} O_j \left(1 - O_j\right)$

Summary of the BP Algorithm

- Adaptation of the weights between output (k) and hidden (j) layers

$$W_{kj}(t+1) = W_{kj}(t) + \Delta W_{kj}(t+1)$$

where $\Delta W_{kj}(t+1) = \eta \delta_k O_j + \alpha \Delta W_{kj}(t)$

$$\delta_k = O_k(1 - O_k)(t_k - O_k)$$

and between the hidden (j) and input (i) layers:

$$W_{ji}(t+1) = W_{ji}(t) + \Delta W_{ji}(t+1)$$

where $\Delta W_{ji}(t+1) = \eta \delta_j O_i + \alpha \Delta W_{ji}(t)$

$$\delta_j = O_j(1 - O_j)\sum_k \delta_k W_{kj}$$

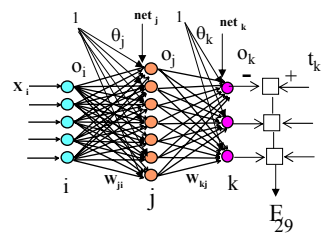
Note that:

$$net_j = \sum_j W_{ji} O_i + \theta_j$$

$$O_j = f(net_j) = \frac{1}{1+e^{-net_j}}$$

$$net_k = \sum_k W_{kj} O_j + \theta_k$$

$$O_k = f(net_k) = \frac{1}{1+e^{-net_k}}$$



Steps of the BP Algorithm

Step 1: Obtain a set of training patterns.

Step 2: Set up neural network model:
No. of Input neurons, Hidden neurons, and Output Neurons.

Step 3: Set learning rate η and momentum rate α

Step 4: Initialize all connection W_{ji}, W_{kj} and bias weights θ_j θ_k to random values.

Step 5: Set minimum error, E_{min}

Step 6: Start training by applying input patterns one at a time and propagate through the layers then calculate total error.

Step 7: Backpropagate error through output and hidden layer and adapt weights.

Step 8: Backpropagate error through hidden and input layer and adapt weights.

Step 9: Check it Error < E_{min}
If not repeat Steps 6-9. If yes stop training.

Past Exam Question 1

The weights of the interconnections have been initialized as shown. If all the neurons are logistic activation function except for the input neurons, layer i, which are linear functions, calculate:

(I) the value of the output O_k of the MLP.

(II) if the target is set to be zero, calculate the first iteration for the change in weights of W_{20},W_{10} and W_{00}. Take learning rate, eta = 0.2.

(III) suppose the output layer neuron k is changed to linear, calculate its output after (II) above has been computed.

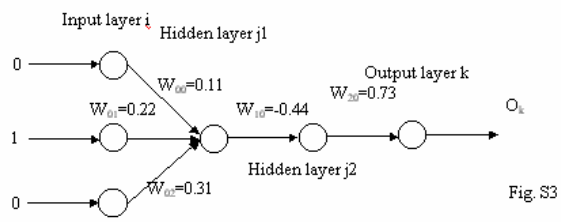


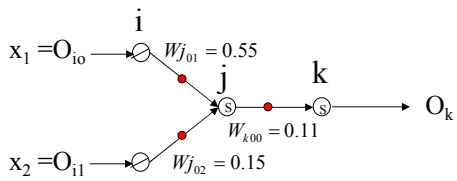
Fig. S3

3.3.4 Solving an XOR Problem

- In this example we use the BP algorithm to solve a 2-bit XOR problem.
- The training patterns of this ANN is the XOR example as given in the table.
- For simplicity, the ANN model has only 4 neurons (2 inputs, 1 hidden and 1 output) and has no bias weights.
- The input neurons have linear functions and the hidden and output neurons have sigmoid functions.
- The weights are initialised randomly.
- We train the ANN by providing the patterns #1 to #4 through an iteration process until the error is minimized.

	Inputs		Target
	O_{i1}	O_{i2}	t_k
Pattern#1 →	0	0	0
Pattern#2 →	0	1	1
Pattern#3 →	1	0	1
Pattern#4 →	1	1	0

- The ANN model with initial weights.



- Training begins when the pattern#1 and its target are provided to the ANN.

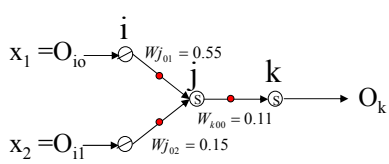
1st pattern: 0, 0 target : 0

<u>On the 1st Iteration</u>:

- At the input of neuron j:

$$net_j = O_{io} x Wj_{01} + O_{i1} . x Wj_{02}$$

$$= 0 \text{ x } 0.55 + 0 \text{ x } 0.15 = 0$$



- At the output of neuron j:

$$O_{j0} = f(net_j) = 1/(1 + \exp(-net_j))$$

$$= 1/ (1 + 1) = 0.5$$

- At the input of neuron k:

$$net_k = O_{jo} x W_{k00}$$

$$= 0.5 \text{ x } 0.11 = 0.055$$

- At the output of neuron k which is the output of the ANN:

$$O_k = f(net_k) = 1/(1+\exp(-net_k))$$

$$= 1/(1+\exp(-0.055))$$

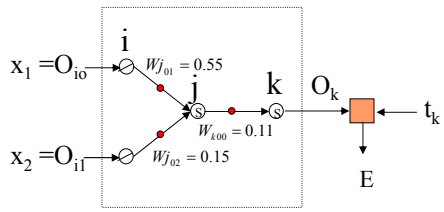
$$= 0.513$$

- Compare this value to the target, we get the error

$$E = (t_k - O_k)$$

Therefore, $E = (0 - 0.513)$

$$= -0.513$$



- This error is now backpropagated through the layers following the error signal equations given as follows:
- Between output (k) and hidden (j) layer

$$\delta_k = O_k(1 - O_k)(t_k - O_k)$$

- Thus, $\delta_k = -0.127$

- Between hidden (j) and input (i) layer :

$$\delta_j = O_j(1 - O_j)\sum \delta_k W_{kj}$$

$$= -0.0035$$

- Now we have calculated the error signal between layers k and j

$$\Delta W_{kj}(t+1) = \eta \delta_k O_j + \alpha \Delta W_{kj}(t)$$

- If we had chosen the learning rate and momentum term as follows :

$$\eta = 0.1 \quad \text{and} \quad \alpha = 0.9$$

and the previous change in weight is 0 and O_{jo}= 0.5 (see Slide 36).

- Then

$$\Delta W_{kj}(t+1) = (0.1 \times -0.127 \times 0.5) + (0.9 \times 0)$$
$$= -0.0064$$

- This is the increment of the weight after the first iteration for the weight between layers k and j.

- Now this change in weight is added to the actual weight as follows

$$W_{kj}(t+1) = W_{kj}(t) + \Delta W_{kj}(t+1)$$
$$= 0.11 + (-0.0064)$$
$$= 0.104$$

and thus the weight between layers k and j has been adapted.

- Similarly for the weights between layers j and i, the adaptation follows

$$\Delta W_{ji}(t+1) = \eta \delta_j O_i + \alpha \Delta W_{ji}(t)$$

- Now this change in weight is added to the actual weight as follows:

$$W_{ji}(t+1) = W_{ji}(t) + \Delta W_{ji}(t+1)$$

  and this is the adapted weight between layers j and i after pattern#1 is seen by the ANN in the first iteration.

- The whole calculation is then repeated for the next pattern (pattern#2 = [0, 1]) with t_k=1.

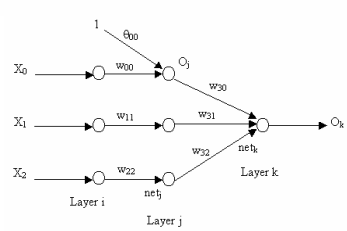
- After all the 4 patterns have been completed the whole process is repeated for pattern#1 again.

Past Exam Question 2

4.

(a). Study the multiplayer "non-fully connected" neural network configuration which is to be trained using the backpropagation algorithm as shown in Fig. 4 and consider the following assumptions/initial parameters:

- the neural network has only 1 bias, θ_{00}.
- all neurons in layers i and k have linear activation functions and all neurons in layer j (hidden neurons) have sigmoid logistic activation functions.
- all the weights between layer i and j are initialized to 0.1 and all the weights between layer j and k and initialized to 0.5, the bias is initialized to be 0.2.
- learning parameter, η =0.1 and ignore momentum term, α.



- Based on the above assumptions:
- (i). First, develop the equations for the output of this neural network (try to write these equations as general as possible).

(4 marks)

- (ii). The algorithm for the adaptation of the weights between the layers i and j is given by where is the error signal between layers k and j. Derive the equation for the adaptation of the weights between the layers k and j.

(6 marks)

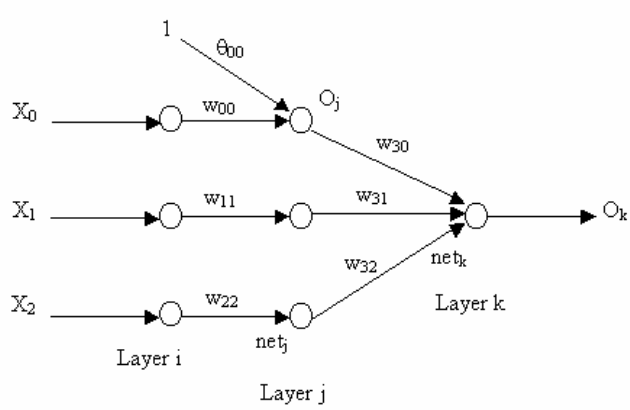
- (iii). Calculate the output O_k of the neural network, when the inputs are as follows:
- $X_0 = 0.5$, $X_1 = 0.7$, and $X_2 = 0.1$ (3 marks)

- (iv). With the above inputs, calculate the new Δw_{30} and also Δw_{11} where the target is 1.

(4 marks)

- (v). Based on (iii) calculate the new bias adaptation, $\Delta\theta_{00}$ (the target is 1).

(3 marks)



3.4 Using the BP for Character Recognition

- Perhaps, the most popular application of the BP algorithm is in pattern classification.
- Pattern classification usually involve image processing techniques.
- One of the earlier applications of the BP is in character recognition/classification.
- Here we discuss the application of the BP algorithm in a project for "Automatic data entry for hand-filled forms" developed at CAIRO.
- We will then use a simple software for understanding character recognition by the BP algorithm.

Design of an Automated Data Entry System for Handwritten Forms

Researchers:
Lim Woan Ning, Marzuki Khalid, Tay Yong Haur and Rubiyah Yusof

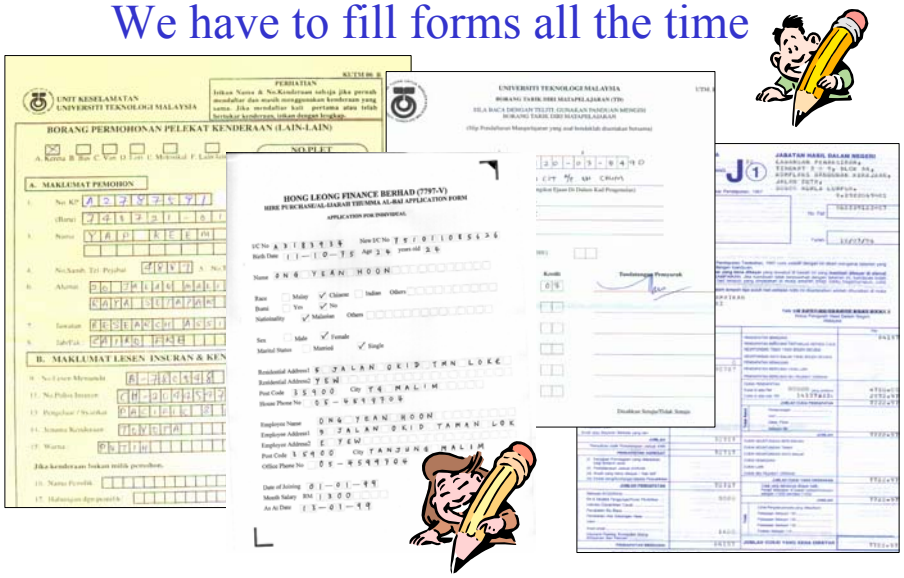
Centre for Artificial Intelligence and Robotics (CAIRO)
Faculty of Electrical Engineering,
Universiti Teknologi Malaysia.

Objectives of Research

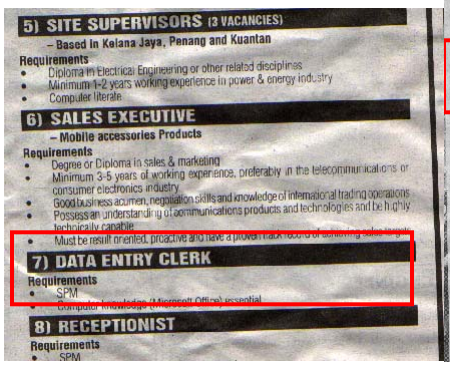
- To develop an automated data entry system that can save laborious manual data entry.
- To design a software that can configure any kind of forms for possible use in government agencies and commercial sectors for high speed data entry.
- To use state of the art technique such as artificial neural networks in handwritten character recognition.
- Towards an IT-based administration system in line with Vision 2020.

Motivation for Research

We have to fill forms all the time



Data Entry work will never be tedious anymore and it will be very cost effective for many companies



5) SITE SUPERVISORS (3 VACANCIES)

– Based in Kelana Jaya, Penang and Kuantan

Requirements
- Diploma in Electrical Engineering or other related disciplines
- Minimum 1-2 years working experience in power & energy industry
- Computer literate

6) SALES EXECUTIVE

– Mobile accessories Products

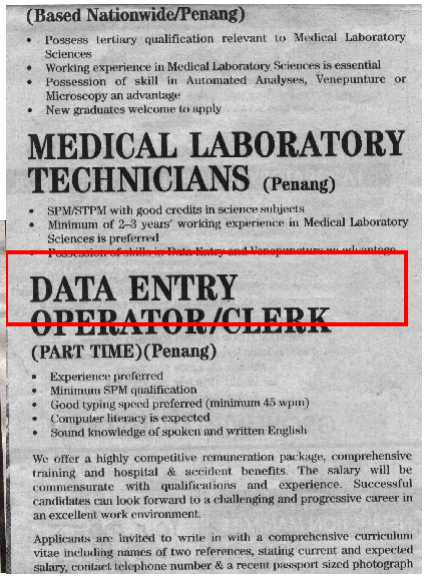
Requirements
- Degree or Diploma in sales & marketing
- Minimum 3-5 years of working experience, preferably in the telecommunications or consumer electronics industry
- Good business acumen, negotiation skills and knowledge of international trading operations
- Possess an understanding of communications products and technologies and be highly technically capable
- Must be result oriented, proactive and have a proven track record of achieving sales targets

7) DATA ENTRY CLERK

Requirements
- SPM
- Computer knowledge (Microsoft Office) essential

8) RECEPTIONIST

Requirements
- SPM



(Based Nationwide/Penang)
- Possess tertiary qualification relevant to Medical Laboratory Sciences
- Working experience in Medical Laboratory Sciences is essential
- Possession of skill in Automated Analyses, Venepunture or Microscopy an advantage
- New graduates welcome to apply

MEDICAL LABORATORY TECHNICIANS (Penang)
- SPM/STPM with good credits in science subjects
- Minimum of 2–3 years' working experience in Medical Laboratory Sciences is preferred
- Possession of skills in Data Entry and Venepunture an advantage

DATA ENTRY OPERATOR/CLERK

(PART TIME)(Penang)
- Experience preferred
- Minimum SPM qualification
- Good typing speed preferred (minimum 45 wpm)
- Computer literacy is expected
- Sound knowledge of spoken and written English

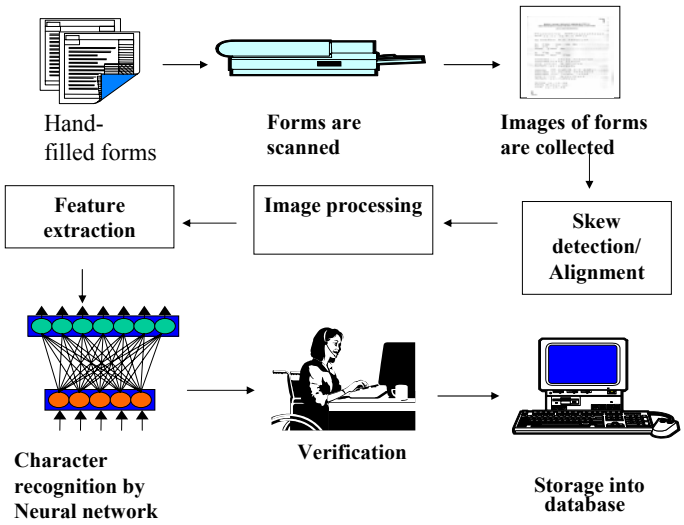
We offer a highly competitive remuneration package, comprehensive training and hospital & accident benefits. The salary will be commensurate with qualifications and experience. Successful candidates can look forward to a challenging and progressive career in an excellent work environment.

Applicants are invited to write in with a comprehensive curriculum vitae including names of two references, stating current and expected salary, contact telephone number & a recent passport sized photograph

Database Systems are used for Handling Information in Large Organizations

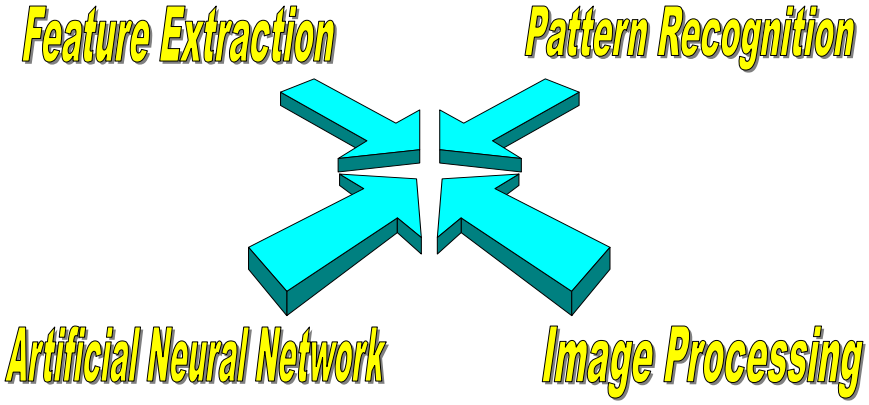
- In many organizations, database systems are used to monitor, execute decisions, and analysis from the information stored.
- Currently, in many organisations humans are needed to key in data from hand-filled forms to be stored into databases.
- Such systems are labor intensive and slow. Thus, the need to automate such process.

Processing stages of our Automated data entry system for hand-filled forms



Research Methodology

A collection of techniques are used in this application



Automated data entry
system for hand-filled forms

For this research, you need the following background:

- Computer systems
- Softwares and operating systems
- Programming knowledge (Visual C++, etc.)
- Management information system (Database)
- Image processing techniques
- Artificial neural networks/Other algorithms
- Related mathematical knowledge
- Related information such as scanner technology, etc.

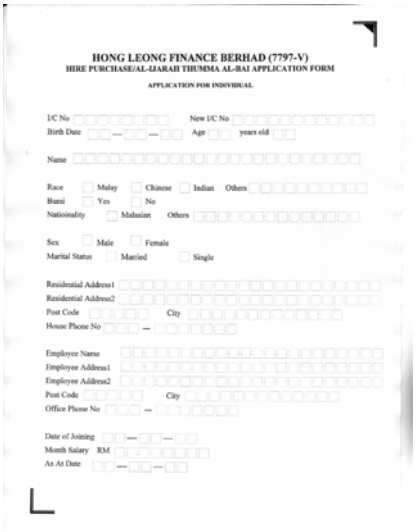
Computer Science

Process Description

STAGE 1

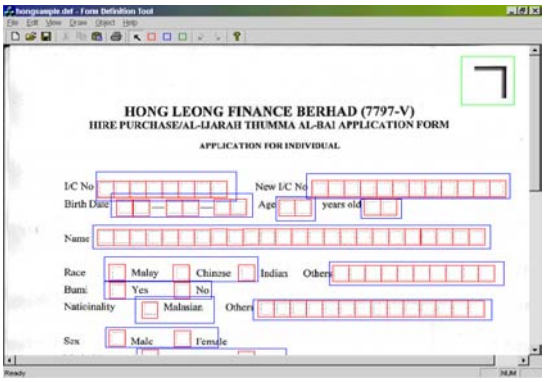
- Forms Scanning
  - Initial stage: The template of the form that is to be used in this system must first be scanned. The template is then stored to be used to identify the location of the characters written in the form. This is done once only for each form.
  - Filled-forms are collected and then scanned.
  - The images are saved in the hard disk to be processed.
  - This can be done automatically with an automatic document feeder (ADF).

Example of a Template



Filled Form

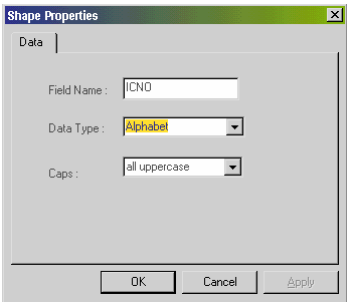
Initial Stage: Template Configuration



- Indicator
- Field
- Character

Configure the fields and characters in the form

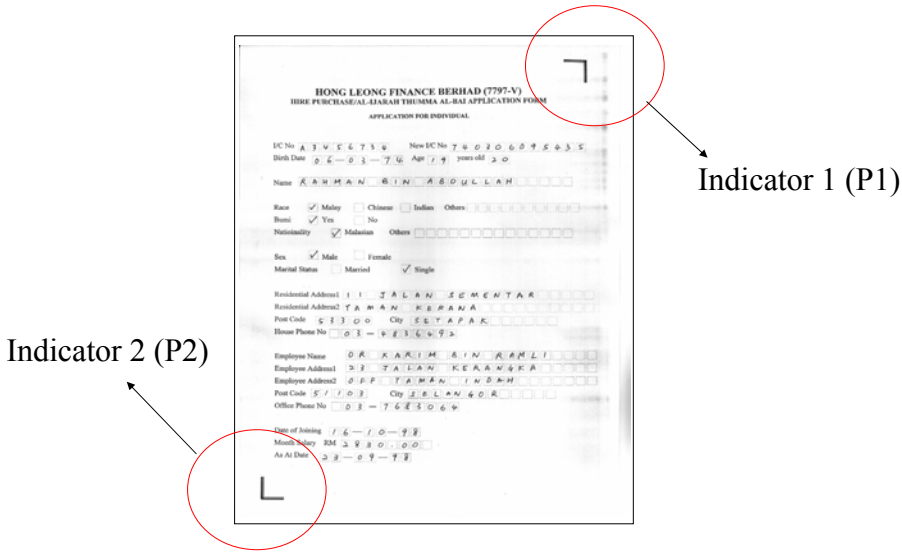
Field's property page



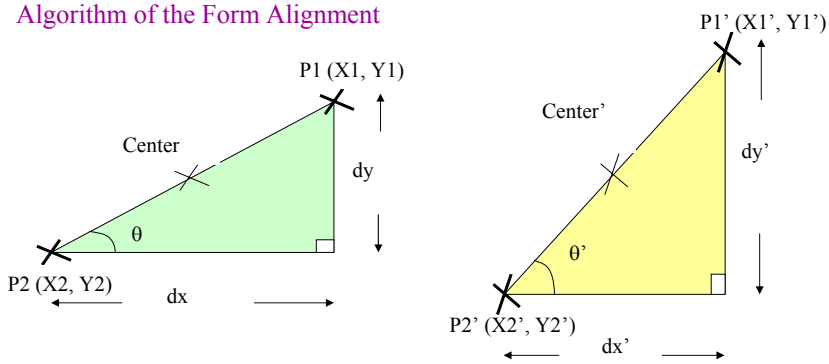
STAGE 2

- Scanned forms need to be align correctly
  - Indicators are used in the forms to detect the images of the scanned forms.
  - The forms are automatically aligned using a skewing algorithm.
  - This is needed to extract the characters from the forms accurately.

Pattern recognition techniques are used to find the form indicators P1 and P2 to calculate the rotation and translation scales.

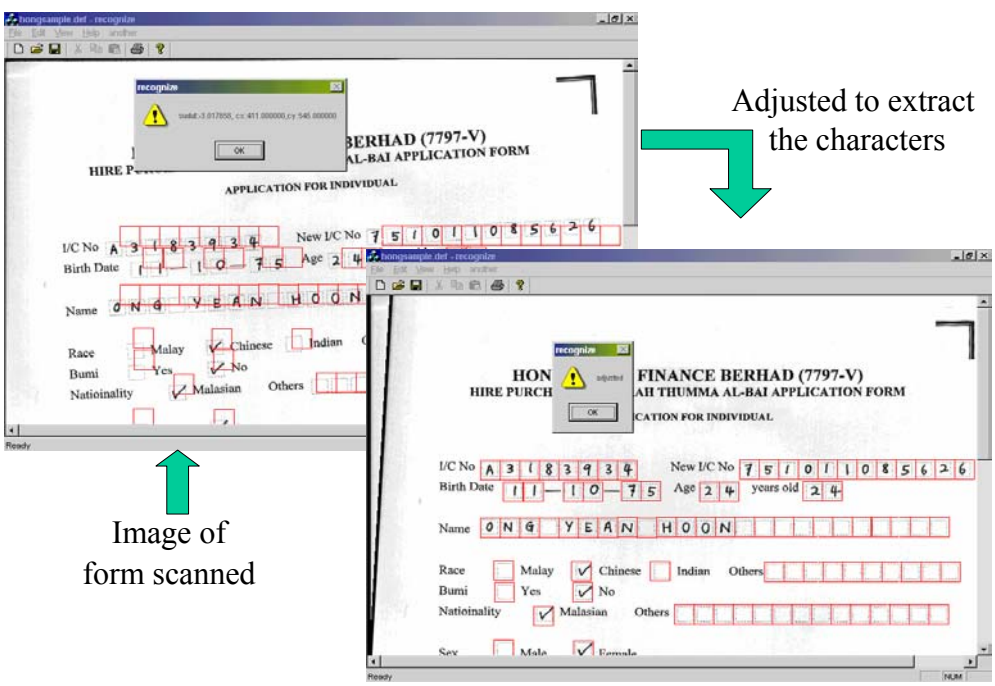


Algorithm of the Form Alignment



P1 P2 = template form indicator
P1' P2' = processed form indicator

<u>Calculation :</u>
Translation = Center - Center'
Rotation = θ - θ'



Recognition Stage

- For each character to be recognized by the neural network, the following processes are required:
  - Each character is extracted from the box
  - Image processing techniques are used as follows:
    - thresholding
    - noise reduction
    - blob analysis
    - normalization
  - Features of the characters are extracted
  - Character Classification: using the BP Algorithm
  - Correction based on the field type

STAGE 3

Original Image

Recognition of the characters

Thresholding, noise reduction, blob analysis, and normalization technqiues are used

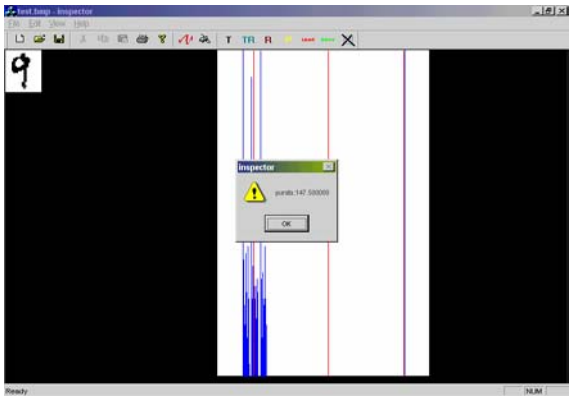
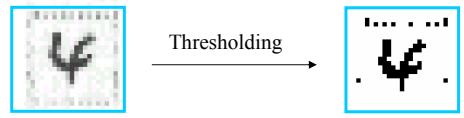


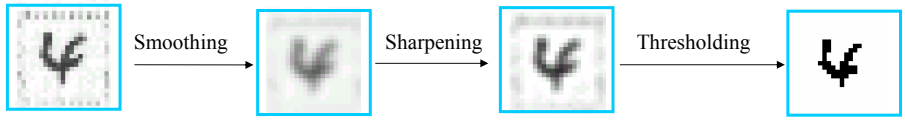
Image Processing Techniques

- Smoothing and sharpening filters are used to reduce the noise of the data
- Otsu algorithm is used for the thresholding



Original Image

Binarized Image With Noise



Original Image

Smoothen Image

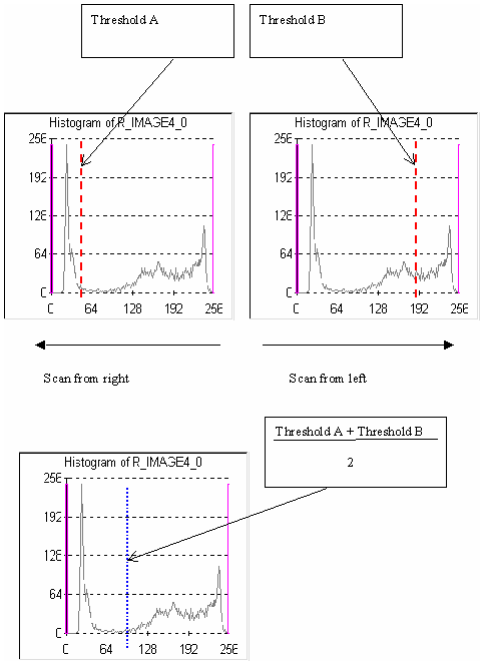
Sharpen Image

Noiseless Binarized Image

Example of Otsu's thresholding technique on histogram to find threshold values.

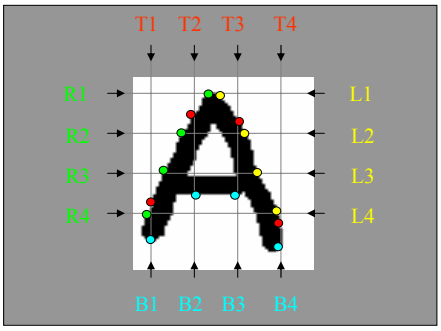
Otsu is a technique that perform threshold values selection from gray level histogram.

Basically, this method is based on probability distribution, zero and first-order of cumulative moments.



Feature Extraction Techniques

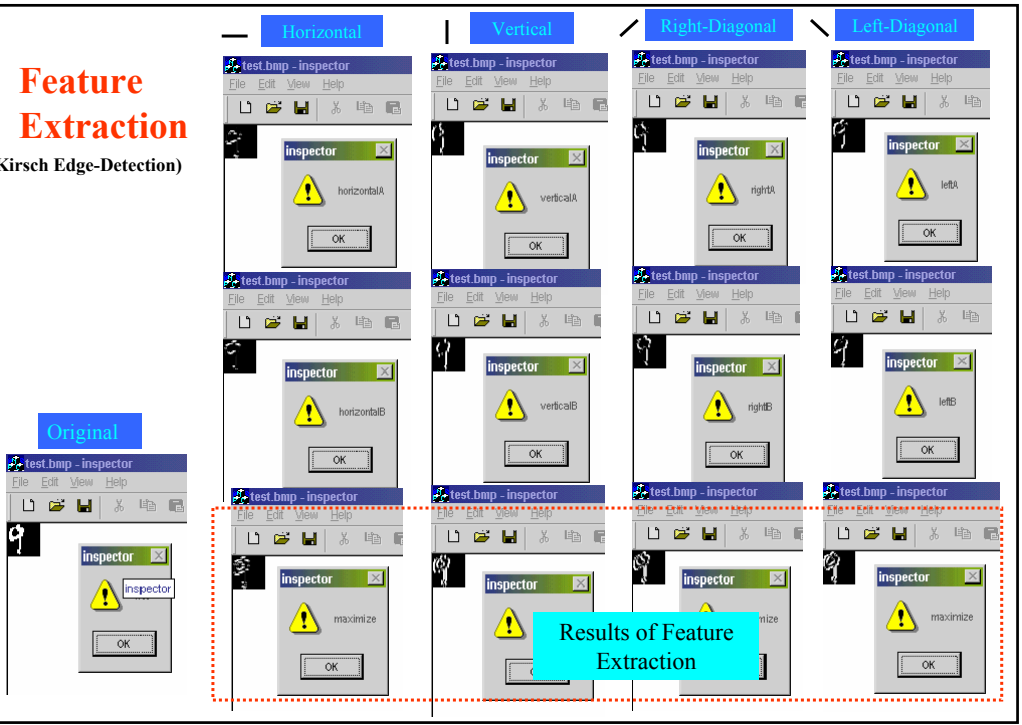
Take the coordinate of the point scan from top, bottom, left and right



- Point of first black pixel scan from right (R)
- Point of first black pixel scan from left (L)
- Point of first black pixel scan from top (T)
- Point of first black pixel scan from bottom (B)

Feature Extraction

(Kirsch Edge-Detection)



Feature Extraction Techniques

Kirsch edge-detection technique is used to extract the features of the characters

Kirsch's four-directional feature masks :
(a) Horizontal (b)Vertical (c) Right-diagonal (d) Left-diagonal.

5	5	5
-3	0	-3
-3	-3	-3

(a)

-3	-3	-3
-3	0	-3
5	5	5

-3	-3	5
-3	0	5
-3	-3	5

(b)

5	-3	-3
5	0	-3
5	-3	-3

-3	5	5
-3	0	5
-3	-3	-3

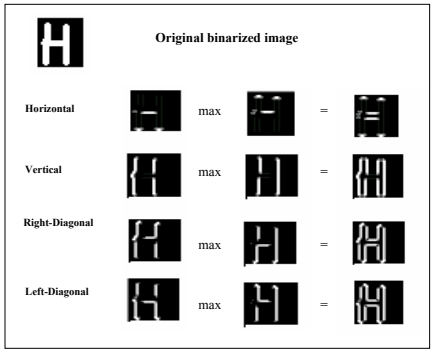
(c)

-3	-3	-3
5	0	-3
5	5	-3

-3	-3	-3
-3	0	5
-3	5	5

(d)

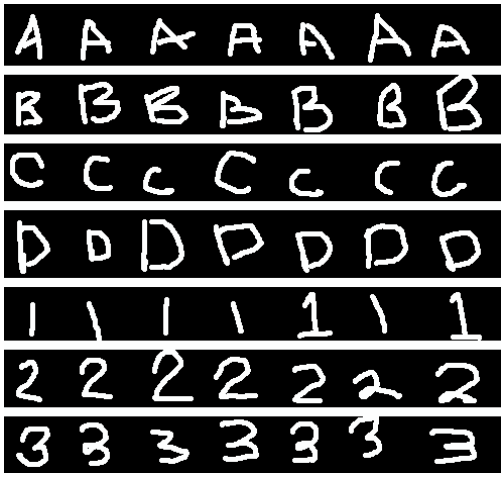
5	5	-3
5	0	-3
-3	-3	-3



Results of character 'H' after applying the Kirsch's mask

Example of the Training Samples

Training Samples



Target

A

B

C

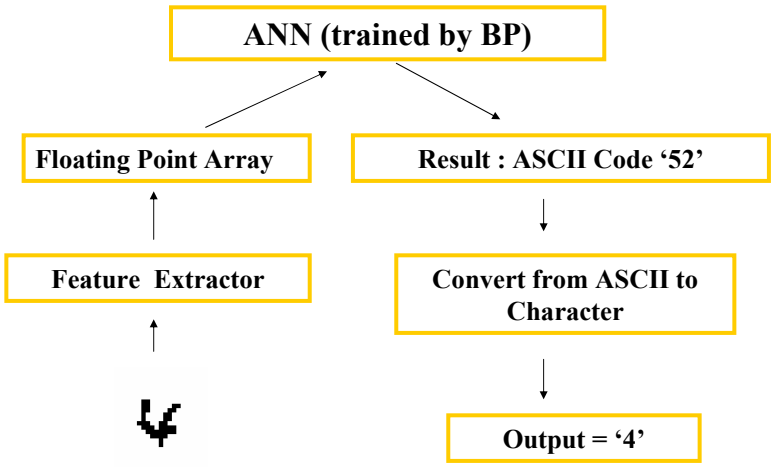
D

1

2

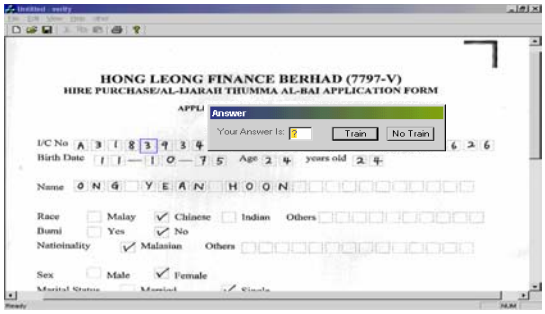
3

Recognition Mode of the ANN



Verification

- After all the three stages of processing have been done where the characters have been identified and stored, verification is still needed.
- This stage requires a clerk to verify all the data.
- For some data which cannot be identified, this is prompted by the system and the clerk can key in the correct character.
- The sample that cannot be recognized can be used to train the neural network further.



STAGE 4

Storage into database

- Once the data has been verified, they are stored automatically into the organization's database for future use.

STAGE 5

```
[FILENAME],ICNO,NEWICNO,BOD,AGE,YEARS,NAME,RACE,OTHERRACE,BUMIPUTRA,NATIONAL
ITY,OTHERNATIONALITY,SEX,MARITALSTATUS,RESADDRESS1,RESADDRESS2,RESPOSTCODE,R
ESCITY,RESTPHONE,EMPNAME,EMPADDRESS1,EMPADDRESS2,EMPPOSTCODE,EMPCITY,EMPPH
ONE,DATEJOINING,SALARY,ATDATE,
[proc1.bmp],A31??434,751011085626,111075,24,24,0NG YEAN HOON        , *,        , *,*,          , *, *,5
JALAN OKLD TMN LOKE ,YEW                ,35900,TG MALIM    , 054599704,0NG YEAN HOON        ,5
JALAN OKID TAMAN LOK,E YEW              ,35900,TANJUNG MALIM, 054599704,0l0199,1300
,130199,
[FILENAME],ICNO,NEWICNO,BOD,AGE,YEARS,NAME,RACE,OTHERRACE,BUMIPUTRA,NATIONAL
ITY,OTHERNATIONALITY,SEX,MARITALSTATUS,RESADDRESS1,RESADDRESS2,RESPOSTCODE,R
ESCITY,RESTPHONE,EMPNAME,EMPADDRESS1,EMPADDRESS2,EMPPOSTCODE,EMPCITY,EMPPH
ONE,DATEJOINING,SALARY,ATDATE,
[proc1.bmp],A31??434,751011085626,111075,24,24,0NG YEAN HOON        , *,        , *,*,          , *, *,
JALAN OKLD TMN LOKE ,YEW                ,35900,TG MALIM    , 054599704,0NG YEAN HOON        ,
JALAN OKID TAMAN LOK,E YEW              ,35900,TANJUNG MALIM, 054599704,0l0199,1300
,130199,
```

ASCII format

Converted to database format

Microsoft Access - [TEST : Table]

ICNO	NEWICNO	AGE	RESPOSTCODE	NAME
A3469434	751011085626	24	66372	ONG TEIK KONG
A2749223	721011035626	27	51000	MOHAMAD AHMAD
A3445434	545101135626	48	33405	YAP JAN NEE
37656445	651011085626	35	23454	TEEM SEE KIAT
A2828334	751011085626	24	53100	RAFIDAH BINTI ALI
A2432355	731011085626	22	43300	KAMARUL BIN AN
A3445444	751011085626	24	54110	SEE KIAT MENG

System requirement:

~ a high-speed scanner

~ a high-end computer

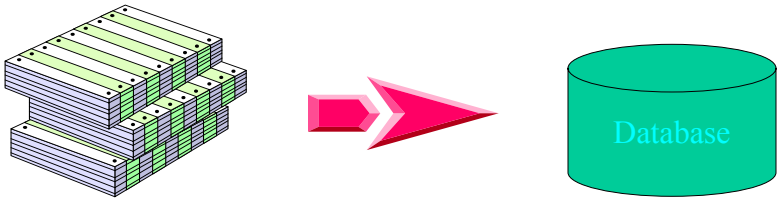
~ our software (Automated data entry engine)

A researcher at CAIRO operating the system.

Scanner with ADF

Outputs of Research

- An intelligent software that can automate data entry process of hand-filled forms
- The software can be used under any Windows 98 OS and can handle several types of scanners
- Commercialization of the whole system as a solution for large organizations (if grant is available)
- A cheaper, lower end product is to be commercialized for smaller organizations/individuals



Conclusion/ Benefits of Research

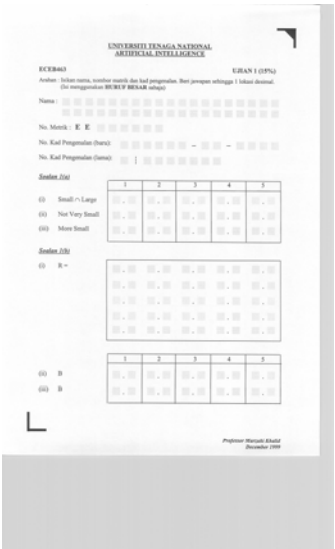
- Development of local expertise in advanced technological areas such as image processing, pattern recognition, intelligent system, in line with the MSC.
- New technology in office automation.
- Generate new skills and knowledge in commercial products.
- Savings on expenses and human resources needed in data entry process. After the economic crisis, human labor will be needed for more demanding jobs.
- Provide substantial increased effectiveness for the data entry process.
- This product can be exported to increase global competitiveness for Malaysia.

Possible Target Applications

Any organization involves in processing large volume of hand-filled forms

- Jabatan Pendaftaran Negara
- Lembaga Hasil Dalam Negeri
- Kementerian Pendidikan
- Jabatan Imigresen
- Jabatan Kastam
- Syarikat Telekom Malaysia
- Tenaga Nasional
- MIDA
- Hospitals
- Banks/Finance Companies
- Survey Research Malaysia
- Govt. of Malaysia National Survey on Population

Application for Marking Exam Papers


UNIVERSITI TENAGA NASIONAL
ARTIFICIAL INTELLIGENCE

ECE8463 UJIAN 1 (15%)

Arahan : Isikan nama, nombor matrik dan kad pengenalan. Beri jawapan sehingga 1 lokasi desimal.
(Sa menggunakan HURUF BESAR sahaja)

Nama :

No. Matrik : E E

No. Kad Pengenalan (baru) : - -

No. Kad Pengenalan (lama) : |

Soalan 1(a)

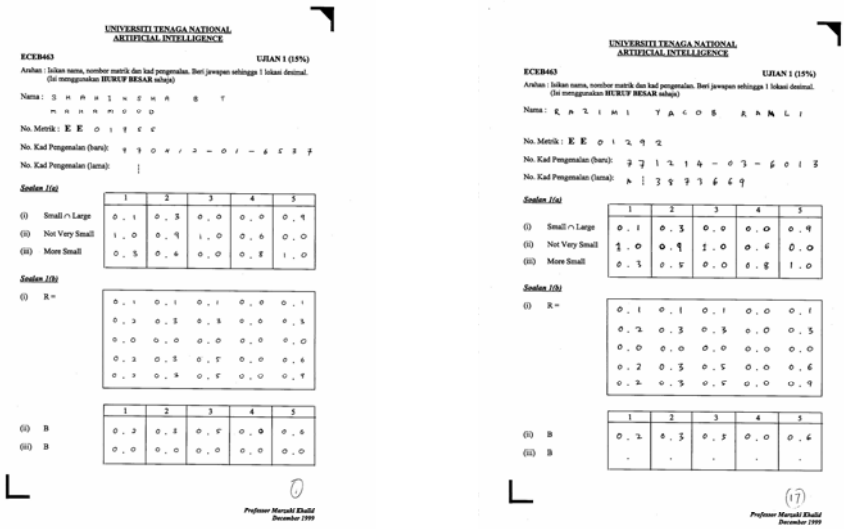
	1	2	3	4	5
(i) Small /\ Large					
(ii) Not Very Small					
(iii) More Small					

Soalan 1(b)

(i) R =

	1	2	3	4	5
(ii) B					
(iii) B					

Professor Marzuki Khalid
December 1999



UNIVERSITI TENAGA NATIONAL
ARTIFICIAL INTELLIGENCE

ECEB463 — UJIAN 1 (15%)

Arahan : Isikan nama, nombor matrik dan kad pengenalan. Beri jawapan sehingga 1 lokasi desimal.
(Sil menggunakan HURUF BESAR sahaja)

Nama : S H A H I N S H A B T M A H M O O D

No. Metrik : E E 0 1 9 5 5

No. Kad Pengenalan (baru): 9 9 0 4 1 2 - 0 1 - 6 5 3 7

No. Kad Pengenalan (lama):

Soalan 1(a)

	1	2	3	4	5
(i) Small ∩ Large	0.1	0.3	0.0	0.0	0.9
(ii) Not Very Small	1.0	0.9	1.0	0.6	0.0
(iii) More Small	0.3	0.6	0.0	0.8	1.0

Soalan 1(b)

(i) R =

0.1	0.1	0.1	0.0	0.1
0.2	0.3	0.3	0.0	0.3
0.0	0.0	0.0	0.0	0.0
0.2	0.3	0.5	0.0	0.6
0.2	0.3	0.5	0.0	0.9

	1	2	3	4	5
(ii) B	0.2	0.3	0.5	0.0	0.6
(iii) B	0.0	0.0	0.0	0.0	0.0

0

Professor Marzuki Khalid
December 1999

UNIVERSITI TENAGA NATIONAL
ARTIFICIAL INTELLIGENCE

ECEB463 — UJIAN 1 (15%)

Arahan : Isikan nama, nombor matrik dan kad pengenalan. Beri jawapan sehingga 1 lokasi desimal.
(Sil menggunakan HURUF BESAR sahaja)

Nama : R A Z I M I Y A C O B R A M L I

No. Metrik : E E 0 1 2 9 2

No. Kad Pengenalan (baru): 7 7 1 2 1 4 - 0 7 - 6 0 1 3

No. Kad Pengenalan (lama): A 1 3 8 7 3 6 6 9

Soalan 1(a)

	1	2	3	4	5
(i) Small ∩ Large	0.1	0.3	0.0	0.0	0.9
(ii) Not Very Small	1.0	0.9	1.0	0.6	0.0
(iii) More Small	0.3	0.5	0.0	0.8	1.0

Soalan 1(b)

(i) R =

0.1	0.1	0.1	0.0	0.1
0.2	0.3	0.3	0.0	0.3
0.0	0.0	0.0	0.0	0.0
0.2	0.3	0.5	0.0	0.6
0.2	0.3	0.5	0.0	0.9

	1	2	3	4	5
(ii) B	0.2	0.3	0.5	0.0	0.6
(iii) B	.	.	.	.	.

(17)

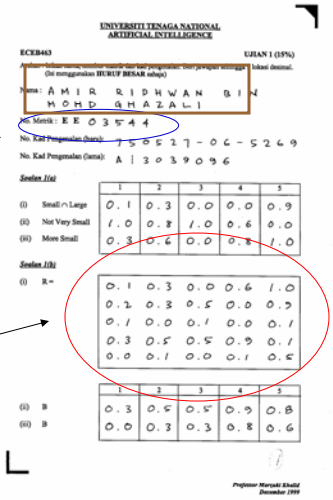
Professor Marzuki Khalid
December 1999

EXAMPLE OF TEST PAPERS FILLED UP BY STUDENTS

AUTOMATIC RECOGNITION OF STUDENTS NAME

METRIC NUMBER

RECOGNITION OF ANSWERS GIVEN BY STUDENTS



UNIVERSITI TENAGA NATIONAL
ARTIFICIAL INTELLIGENCE

ECEB463 — UJIAN 1 (15%)

Arahan : Isikan nama, nombor matrik dan kad pengenalan. Beri jawapan sehingga 1 lokasi desimal.
(Sil menggunakan HURUF BESAR sahaja)

Nama : A M I R R I D H W A N B I N M O H D G H A Z A L I

No. Metrik : E E 0 3 5 4 4

No. Kad Pengenalan (baru): 7 5 0 6 2 7 - 0 6 - 5 2 6 9

No. Kad Pengenalan (lama): A 1 3 0 3 9 0 3 6

Soalan 1(a)

	1	2	3	4	5
(i) Small ∩ Large	0.1	0.3	0.0	0.0	0.9
(ii) Not Very Small	1.0	0.8	1.0	0.6	0.0
(iii) More Small	0.3	0.6	0.0	0.8	1.0

Soalan 1(b)

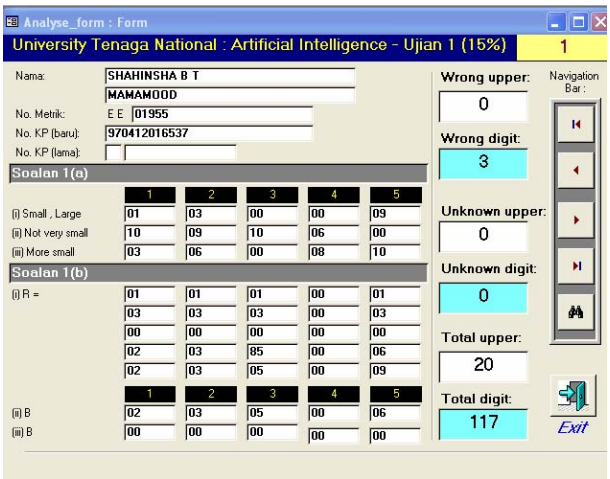
(i) R =

0.1	0.3	0.0	0.6	1.0
0.2	0.3	0.5	0.0	0.3
0.1	0.0	0.1	0.0	0.1
0.3	0.5	0.5	0.9	0.1
0.0	0.1	0.0	0.1	0.5

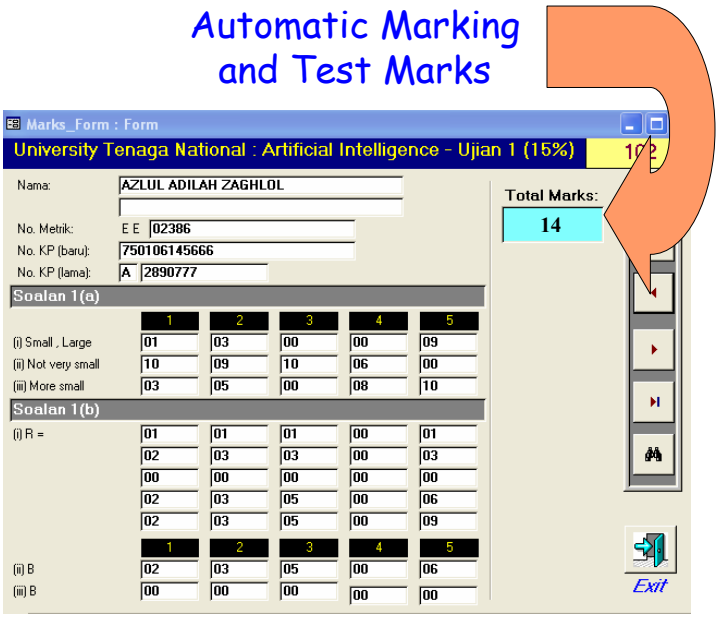
	1	2	3	4	5
(ii) B	0.3	0.5	0.5	0.9	0.8
(iii) B	0.0	0.3	0.3	0.8	0.6

Professor Marzuki Khalid
December 1999

RECOGNITION RESULTS



Automatic Marking and Test Marks



3.4 Summary of Module 3

- We have discussed about multilayer neural networks which are basically feedforward neuro-models consisting of hidden neurons.
- We have discussed about the logistic activation function which is also known as the sigmoid function. This functions are used in the hidden and output neurons of the BP algorithm.
- The backpropagation algorithm has been extensively derived.
- We have shown how the BP can be used to solve a simple XOR problem.
- We have discussed a practical application of the BP algorithm for solving handwritten characters.