INTELLIGENT AGENT FOR ELECTRONIC COMMERCE

by

LAI SIEW CHENG

WEK 98298

A thesis presented to the Faculty of Computer Science and Information Technology of University of Malaya in partial fulfillment of the Requirements for the Degree of

BACHELOR OF COMPUTER SCIENCE

UNIVERSITY OF MALAYA

FEBRUARY 2001

University of Malaya

TABLE OF CONTENTS

University of Malaya

TABLE OF CONTENTS (CONTINUED)

University of Malaya

TABLE OF CONTENTS (CONTINUED)

University of Malaya

ACKNOWLEDGEMENT

I would like to take this opportunity to express my gratitude to a few persons who have contributed their help to me. First, I like to express my gratitude to my supervisor, Mr. Woo Chaw Seng, who has guided us throughout this project and always give his valuable opinions to me. I also owe my gratitude and appreciation to both my parents because they always supported me in everything I do. Gratitude is also due to my partner, Miss Yap Hui Sun, because without her this project cannot be completed. Finally, I would like to thank all of my friends who have provided their help to me in every ways.

University of Malaya

LIST OF FIGURES

Figure *Page*

University of Malaya

LIST OF TABLES

University of Malaya

ABSTRACT

The objective of this project is to develop a system that can assists a user to make decision in online transactions for residential houses. In order to achieve this objective, an intelligent agent will be built where it will help the user to find relevant information of the houses based on they requirement. Furthermore, a prediction tool will also be developed to predict the price of the required house on the market. Artificial neural network will be used here, where the LVQ network is used to build the system. The neural network will be implemented in the filtering module and the prediction module of the system. Matlab has been chosen as the software to develop the network since it contained the neural network toolbox and the database toolbox to import and export data to the Matlab environment. After implementation, the system is supposed to perform two tasks and that are to provide user with relevant information of the required house and the predicted price of the house in the current market.

The system has been built, and the implementation processes where, how the neural network and the interfaces are developed are described in details. After that, the system is evaluated to ensure that it has met the goal set earlier. Here, the system strengths and limitations will be discussed. Finally, a conclusion of the system on the whole is presented at the end of the report.

University of Malaya

CHAPTER I

INTRODUCTION

This introductory chapter gives a description or purpose of the project and problems to be solved. The significance and rationale of the project will be discussed here. Furthermore, the system functions, limitations and its assumptions will also be dwelt into later in this chapter.

1.1 Project Definition

Since the creation of search engine, the task of finding information on the Internet has become easier. Unfortunately, there are still some problems, which this tool cannot deal with. Most of the time when we find information using a search engine, we are bound to get some irrelevant information or in some cases information not related at all to what the user demanded. The search also returned a large amount of URL related to the query and the user has to browse every one of them to get what they want. This task is very time-consuming and they may not find what they want. To overcome this problem, an intelligent agent is built to retrieve relevant information and display them to the user. This project is aim at building an intelligent agent that collects information from the web to assist users in decision-making.

The domain of this project is on the houses. This agent is used to help the user to retrieve only relevant information required by the user from the Internet. Beside this, a prediction

University of Malaya

tool will be built as to assist user in decision-making. Artificial neural network approach will be used in developing this system in which the LVQ network is chosen as the neural network to be used.

1.2 Objective

The main objective of this project is to develop an agent that is intelligent enough to collect relevant information from the World Wide Web based on the user requirements. This yet to be built agent will be concentrated on the domain of houses, which are available online. The housing development in Malaysia is progressing very fast especially for the last few years. Because of this, a lot of housing developers and property agents had put their web site on the Internet for the purpose of providing information about the houses that are for sale to the users.

Houses are built everywhere through out the country and it also come in various sizes and types. In order to search for a particular kind of houses, the user have to consider the location, types, sizes, prices, and so on. This proved to be a time-consuming task especially if the user searches using search engine on the Internet because it will only returned the URL of the sites. They may have to browse through many web sites to get what they want and in some cases the sites provided is not relevant at all to what the user required. Even though the sites are relevant, user still has to go to these sites to compare the prices, location and so on.

This problem can be solved using an intelligent agent to help the user searches for what they want. This agent will represent the user to do the searching on the Internet and it will display the result based on the houses location, prices, sizes, and so on which are specify by the user. User just has to submit the query and the agent will help them to find what they want. When the results are returned, the agent will display the list of the related URL based on the user requirements. When the agent is working, user can do some more enjoyable task while waiting for the result. The query entered by user will also be used to predict the price of the specified house. The rationale of this project is to build an intelligent agent for the above purpose.

University of Malaya

1.3 Scope

The function of this system is to provide an agent to help the users to search for a house on the web based on their requirement. Basically, the agent will help the user in decision-making based on the predicted prices of the products.

This system has its own limitations. The domain covered in this system is on houses. The searching will be conducted on the houses available only in Malaysia.

The output of the system will display the following information of the products:

1. Information of the houses such as:
   - Property type
   - Price
   - Area
   - Related URL
2. Predicted market prices of the houses.

University of Malaya

1.4 Project Schedule

The Gantt chart below shows the schedule of this project.

No	Task	Start Date	End Date	Duration
1	Literature Review	6/6/00	7/14/00	29d
2	System Analysis	7/3/00	7/20/00	14d
3	Methodology	7/20/00	8/1/00	9d
4	Designing System	7/31/00	8/10/00	9d
5	Documentation of 3181 Proje	8/11/00	9/1/00	16d
6	Semester Exam	9/4/00	9/25/00	16d
7	Learn Software Used	9/26/00	10/27/00	24d
8	Develop System Prototype	10/30/00	11/17/00	15d
9	Coding	11/20/00	12/26/00	27d
10	Integrate System	12/27/00	1/1/01	4d
11	Testing	1/2/01	1/12/01	9d
12	System Documentation	1/15/01	1/22/01	6d

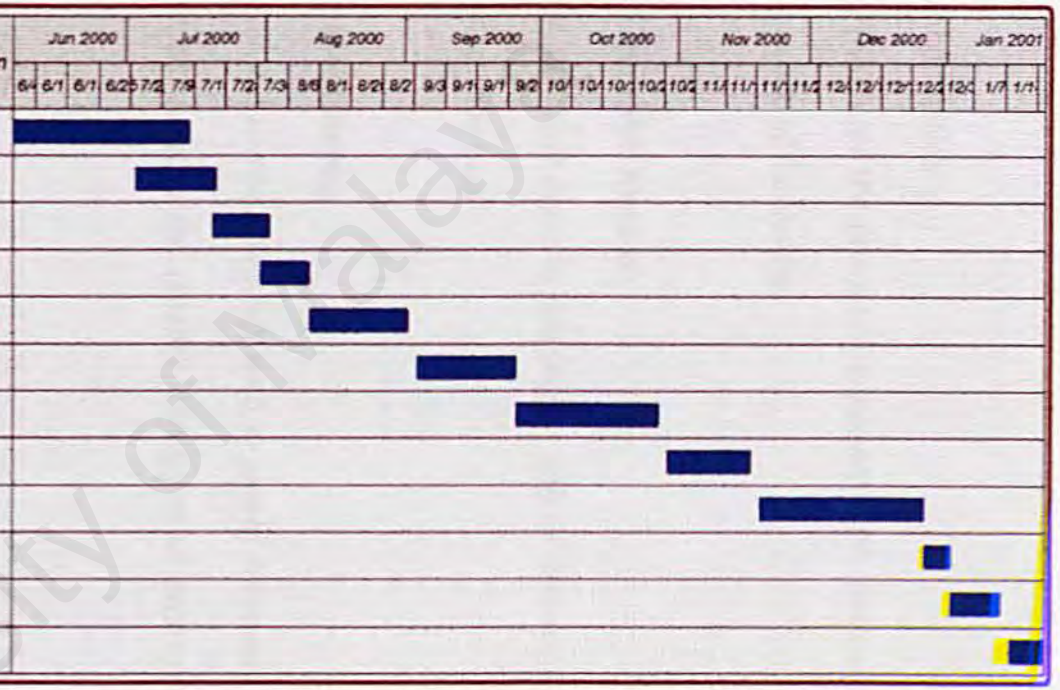


Figure 1.1 Gantt Chart For The Project

University of Malaya

1.5 Chapter Overview

Chapter I: Introduction

This chapter introduces the project in a general view and the rationale of the project. It also states the scope of the system.

Chapter II: Literature Review

In this chapter, the areas cover in this project will be discussed. Furthermore, the existing system is also covered here.

Chapter III: Methodology

For this chapter, the methods that are used to gather data and information for this project will be reviewed. Also in the chapter, the types of prototyping and software use are discussed.

Chapter IV: System Design

The system design will be presented in chapter IV where the system data flow diagram is shown here. The architecture of the neural network will also be shown here.

Chapter V: System Implementation And Testing

This chapter will discuss about the steps and methods taken to implement the system that was designed in previous chapter. After that, the discussion of the system testing will follow in which a few kinds of testing the system is presented.

University of Malaya

Chapter VI: System Evaluation And Conclusion

In Chapter VI, the system built will be evaluated. The criteria that will be evaluated are system strengths, limitations and future enhancement. For the conclusion part, the problems encountered during the process of the entire system and knowledge gained will be stated. The overall conclusions of the system will be the final part in this chapter.

The following chapter will discuss about the areas related to this project and the analysis of some existing system will also be covered.

University of Malaya

CHAPTER II

LITERATURE REVIEW

This chapter will discuss about the meaning and overview of some areas covered in the project literally. The major areas that will be covered are as follow:

- Artificial Intelligence's Approaches
- Artificial Neural Network
- Intelligent Agent
- Electronic Commerce and The Internet

2.1 Artificial Intelligence's Approaches

Artificial Intelligence (AI) is the studies of how to make computers do things, which, at the moment, people do better [1].

There are many problem-solving approaches in artificial intelligence and some of the popular approaches are heuristic (Rule-of-Thumb), fuzzy logic, case-based reasoning, genetic algorithm, artificial neural network, etc.

Three approaches can be applied to this project and that includes Case-Based Reasoning (CBR), Genetic Algorithm (GA) and Artificial Neural Network (ANN). Each of these approaches will be introduced here to give a general idea about it.

ANN will be discussed in section 2.2.

University of Malaya

2.1.1 Case-Based Reasoning

Computer systems that solve new problems by analogy with old ones are often called CBR. A CBR system draws its power from a large case library, rather than from a set of first principle [1]. In other words, the CBR system expertise is embodied in a library of past cases, rather than being encoded in classical rules. Each case typically contains a description of the problem, plus a solution or the outcome. When a problem is solved, it is automatically stored in the case library.

All case-based reasoning methods have in common the following process:

- Retrieve the most similar cases comparing the case to the library of past cases.
- Reuse the retrieved case to try to solve the current problem.
- Revise and adapt the proposed solution if necessary.
- Retain the final solution as part of a new case.

Since the appearance of CBR, it has been applied in a wide range of domains such as diagnosis, help desk, assessment, and decision support, design and so on. When using CBR, the need for knowledge acquisition can be limited to establishing how to characterize cases. CBR allows the case-base to be developed incrementally, while maintenance of the case library is relatively easy and can be carried out by domain experts.

University of Malaya

2.1.2 Genetic Algorithms

Genetic algorithms are search procedures that use the mechanics of natural selection and natural genetics. The genetic algorithm, first developed by John H. Holland in the 1960's, allows computers to solve difficult problems. It uses evolutionary techniques, based on function optimization and artificial intelligence, to develop a solution. The basic operation of a genetic algorithm is quite simple to understand [5].

Most symbolic AI system is very static. Usually, these systems are designed to solve one specific problem. If the given problem were somehow changed, these systems could have a hard time adapting. The algorithm that would have originally led to the right solution could be either incorrect or less efficient. Genetic algorithms were created to combat these problems. These systems are based on natural biological evolution. Their evolution-centered architecture allows them to perform searches much more efficiently then most symbolic AI systems.

A GA's first step is to generate a large set of random possible solutions sets to a given problem (this is analogous to chromosomes). It then evaluates each of those solutions, and then decides on a "fitness level" for each solution set (or chromosomes). The more "fitter" a certain chromosome is, the more likely it is to mate and reproduce. The theory behind this is based on natural selection. In biology, species that are more "fit" or adapted to their environment have a higher chance of surviving, and thus reproducing. When each pair of solution sets (chromosomes) reproduce, their offspring becomes a combination of the parent solution sets. The composition of the mother solution set is mixed in with the

University of Malaya

father solution set. The solution sets from the next generation then mutate and reproduce in the same fashion, just as their parents had. This cycle continues on and on, until one chromosome is found to have the desired solution set. This chromosome would have the most desirable fitness level.

University of Malaya

2.2 Artificial Neural Network

In this section, the general understanding and history of ANN will be discussed. Then it will be classified into a few categories and the how the ANN works will also be shown. Finally, the back propagation learning methods and the Learning Vector Quantization network will be dealt with.

2.2.1 Introduction

The fundamental processing element of a neural network is a neuron. This building block of human awareness encompasses a few general capabilities. Much is still unknown about how the brain trains itself to process information, so theories abound. In the human brain, a typical neuron collects signals from others through a host of fine structures called dendrites. The neuron sends out spikes of electrical activity through a long, thin stand known as an axon, which splits into thousands of branches.

At the end of each branch, a structure called a synapse converts the activity from the axon into electrical effects that inhibit or excite activity from the axon into electrical effects that inhibit or excite activity in the connected neurons. When a neuron receives excitatory input that is sufficiently large compared with its inhibitory input, it sends a spike of electrical activity down its axon. Learning occurs by changing the effectiveness of the synapses so that the influence of one neuron on another changes. Figure 2.1 below show a biological neuron in a human brain.

University of Malaya

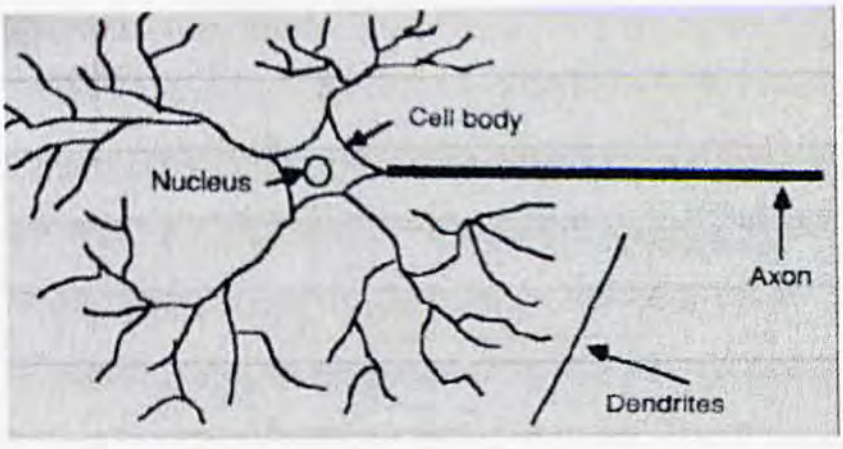


Figure 2.1 A Biological Neuron

Artificial neural networks are parallel computational models comprised of densely interconnected adaptive processing units. These networks are fine-grained parallel implementations of nonlinear static or dynamic systems. A very important feature of these networks is their adaptive nature, where "learning by example" replaces "programming " in solving problem [2].

The term "neural network" describes a class of models, which appear under different names in the literature: neural networks, neural computation, artificial neural systems, connectionist models and parallel-distributed models [3].

University of Malaya

This artificial neural network is comprised of various neurons connected by its weight. These neurons are modeled after the biological neuron shown above. Figure 2.2 show the architecture of a basic artificial neuron.

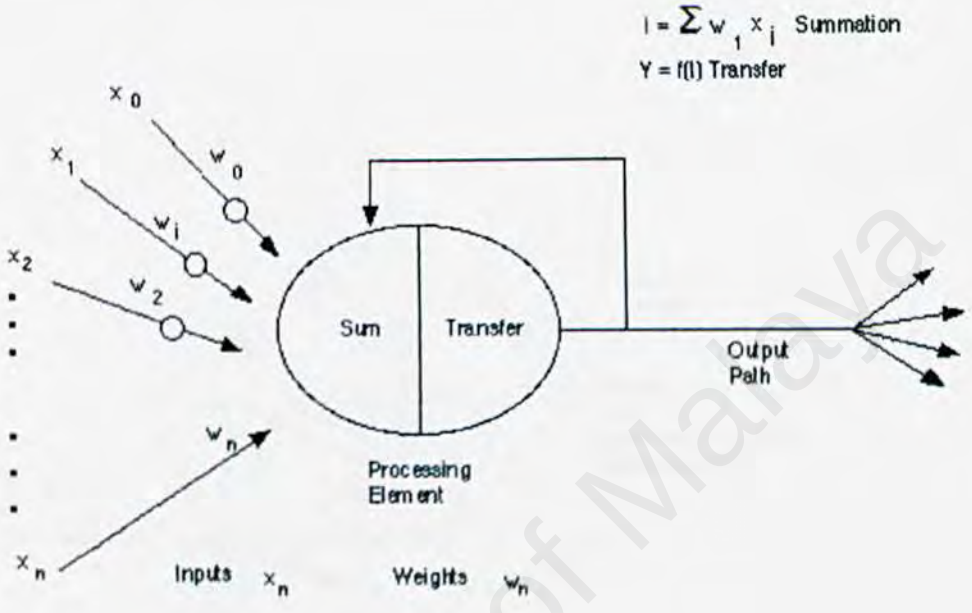


Figure 2.2 A Basic Artificial Neuron.

An artificial neuron is a basic processing element of a neural network. Refer to the above figure the various inputs to the network is represented by X (n). This input is analog to the dendrites of a biological neuron. Each of this input is multiplied by a connection weight W (n). This product will be summed together and send to a transfer function to generate a result. There are many different transfer functions, which can be used such as hard limit, pure linear, log-sigmoid, etc.

University of Malaya

The neural network can consist of a lot of layers. A single layer is combined of one or more artificial neurons. Figure 2.3 below shown a simple neural network with three layers namely input layer, hidden layer and output layer.

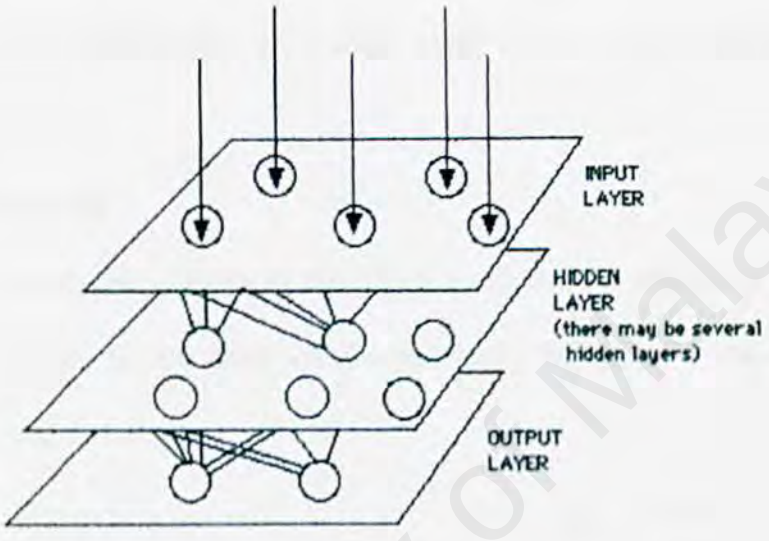


Figure 2.3 A Simple Neural Network Diagram

2.2.2 Classification

There are many kinds of neural network based on their learning algorithms and network topologies. Basically, there are 2 major learning algorithms in ANN, which are supervised learning and unsupervised learning. Network topologies in ANN are divided in two categories, which consist of feed forward and feedback network.

Each of this learning methods and network topologies will be discussed below:

University of Malaya

Learning Algorithm

One of the most significant attribute of ANN is its to learn by interacting with its environment or with information source. Learning in a neural network is normally accomplished through an adaptive procedure, known as a learning rule or algorithm, whereby the weights of the network are incrementally adjusted so as to improve a predefined performance measure over time [2]. Currently, learning rules can be divided into three main types. These three types are supervised, unsupervised and reinforced.

1) Supervised Learning

In supervised learning, the network is trained on a training set consisting of vector pairs. One vector is applied to the input of the network; the other is used as a "target" representing the desired output.

2) Unsupervised Learning

Unsupervised learning, sometimes called self-organization requires only input vectors to train the network. During the training process the network weights are adjusted so that similar inputs produce similar output. This is accomplished by the training algorithm that extracts statistical regularities from the training set, representing them as the values of network weights [4].

University of Malaya

3) Reinforced Learning

Reinforcing learning is unlike the other two learning methods mentioned above, where network is trained with sample outputs. Instead reinforced learning use punishment and reward to train the network. This method is similar to human training an animal to perform some specific task by giving it reward or punishment. First, the network is presented with a sample input and the network computes what it thinks should be the sample output. Then the network is supplied with a real valued judgment from the teacher. If the value is positive, then the network will receives a reward and vice versa. After this, the weights are adjusted and the process continues.

Network Topology

There are two major kinds of network topology, which are feed forward and feedback neural network.

1) Feed forward Network

In a feed forward neural network, the connections between units do not form cycles. Feed forward neural networks usually produce a response to an input quickly. Most feed forward neural networks can be trained using a wide variety of efficient conventional numerical methods in addition to algorithms invented by neural network researchers.

University of Malaya

2) Feedback Network

In a feedback or recurrent neural network, there are cycles in the connections. In some feedback neural networks, each time an input is presented, the neural network must iterate for a potentially long time before it produces a response. Feed back neural networks are usually more difficult to train than feed forward neural networks.

The two learning methods are closely connected with a certain network topology. Below is the list of the names of neural networks based on the order of learning method and network's topology.

The classification of neural networks according to learning rules and topologies.

1. **Supervised learning (i.e. with a "teacher"):**

   i. Feedback Networks:

   a). Brain-State-in-a-Box (BSB)

   b). Fuzzy Congitive Map (FCM)

   c). Boltzmann Machine (BM)

   d). Mean Field Annealing (MFT)

   e). Recurrent Cascade Correlation (RCC)

   f). Back propagation through time (BPTT)

   g). Real-time recurrent learning (RTRL)

   h). Recurrent Extended Kalman Filter (EKF)

University of Malaya

ii. Feed forward-only Networks:

a). Perceptron

b). Adaline, Madaline

c). Back propagation (BP)

d). Cauchy Machine (CM)

e). Adaptive Heuristic Critic (AHC)

f). Time Delay Neural Network (TDNN)

g). Associative Reward Penalty (ARP)

h). Avalanche Matched Filter (AMF)

i). Backpercolation (Perc)

j). Artmap

k). Adaptive Logic Network (ALN)

l). Cascade Correlation (CasCor)

m). Extended Kalman Filter (EKF)

n). Learning Vector Quantization (LVQ)

o). Probabilistic Neural Network (PNN)

p). General Regression Neural Network (GRNN)

2. Unsupervised learning (i.e. without a "teacher"):

i. Feedback Networks:

a). Additive Grossberg (AG)

b). Shunting Grossberg (SG)

University of Malaya

c). Binary Adaptive Resonance Theory (ART1)

d). Analog Adaptive Resonance Theory (ART2, ART2a)

e). Discrete Hopfield (DH)

f). Continuous Hopfield (CH)

g). Discrete Bidirectional Associative Memory (BAM)

h). Temporal Associative Memory (TAM)

i). Adaptive Bidirectional Associative Memory (ABAM)

j). Kohonen Self-organizing Map/Topology-preserving map (SOM/TPM)

k). Competitive learning

ii. Feed forward-only Nets:

a). Learning Matrix (LM)

b). Driver-Reinforcement Learning (DR)

c). Linear Associative Memory (LAM)

d). Optimal Linear Associative Memory (OLAM)

e). Sparse Distributed Associative Memory (SDM)

f). Fuzzy Associative Memory (FAM)

g). Counterprogation (CPN)

University of Malaya

2.2.3 How ANN works

Artificial neural networks are made up of many artificial neurons connected together by some weights. In an artificial neuron there are actually seven important components that make an ANN functional. The seven components are weighting factors, summation function, transfer function, scaling and limiting, output function, error function and learning function.

1) Weighting Factors

A neuron usually has many input vectors and each of these inputs has its own relative weights. The function of the weight here is to give the impact input has on the summation function. The weight in the neuron has the same function of the synaptic strength in a biological neuron. There is a measure of input's connection strength. In some cases, some input may be important than the others, thus the weight of these important inputs is strengthened to provide a response to the network. In order to get an appropriate response, the network can be trained by adjusting its weight. The weighting factors is an important feature in ANN where a network is trained to perform some specific tasks by changing the weight.

2) Summation Function

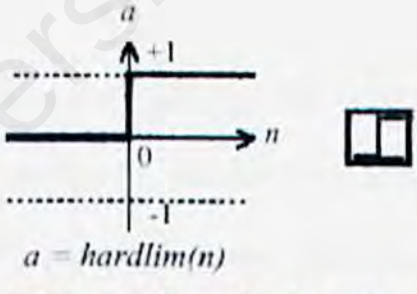
The weight and input of a neuron can be represented by (w1, w2,....., wn) and (i1, i2, ..., i3) respectively. The total input signal of the neuron is the product of these two components where Input1 = i1*w1, Input2 = i2*w2 and so on. Total input is the sum of all these input. Total input = Input1 + Input2 + ... + Inputn. Other than summation,

University of Malaya

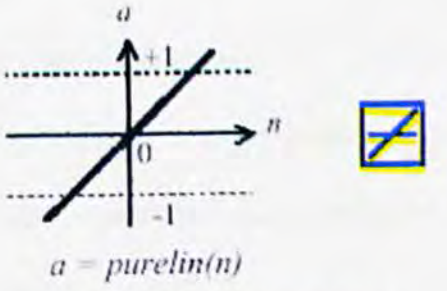
these each input can be combined in many others ways before it is passed to the transfer function. The combination is depended of the network architecture and paradigm.

3) Transfer Function

The result of the summation function will be passed to the transfer function to produce an output. In the transfer function, the summation is compared to some threshold value to determine its output. The threshold or transfer function is normally non-linear but linear function was used in early works, which make it not very useful because input is just proportional to the input. Shown in figure 2.4 are some of the transfer functions used in ANN. Noise may be added to the transfer function to bring out more brain-like transfer function because noise is the temperature of the neuron and thus can affect the ability to think.



Hard Limit Transfer Function



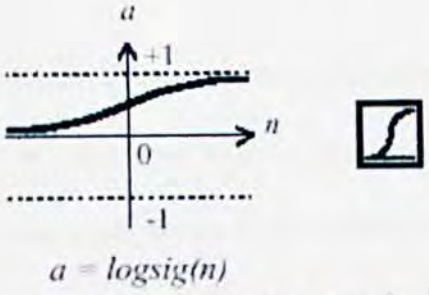


Figure 2.4 Example of Transfer Functions

4) Scaling and Limiting

This is just an additional process to scale and limit the result from the transfer function. The scaling is just a process to multiply the transfer value and add an offset to it. For the limiting process, the scaled result is limited to some upper or lower value.

5) Output Function

Every output can be can be connected to many other neurons. Usually, the output value is the direct value from the transfer function but in some cases the output is modified to incorporate competition among neighboring processing elements.

University of Malaya

6) Error Function and Back-Propagated Value

The difference between the current output and the desired output is calculated and this raw error is transformed by the error function to match particular network architectures.

2.2.4 Back propagation

Back propagation is a form of supervised learning for multi-layer nets, also known as the generalized delta rule. Error data at the output layer is "back propagated" back to earlier ones, allowing incoming weights to these layers to be updated. It is most often used as training algorithm in current neural network applications [3].

A back propagation network starts with a set of random weights and it adjusts its weight every time it sees an input-output pair. Each of the input-output pair will go through two steps: a forward pass and a backward pass. In the forward pass, the network is presented with sample input and let the activation flows until it reached the output layer. Then in the backward pass, the network actual output is compared with the target output and in the meantime error is computes from the differences between these two outputs. When the error is estimated, the weights connected to the output unit are adjusted to reduce error. Then the error that is estimated is used to derive error estimates in the hidden layer. Finally, errors are propagated back to the connection stemming from the input layer. The training process can be summarized as follow:

University of Malaya

1) Initialize weight (randomly).
2) Repeat for each pairs in the training set the following steps:
   i) Apply input vectors. (Forward pass)
   ii) Calculate output vector.
   iii) Calculate the error at output layer.
   iv) Adjust weight layer by layer to reduce the error (backward pass) until performance is satisfactory.

2.2.5 Learning Vector Quantization (LVQ) network

Learning vector quantization (LVQ) is a method for training competitive layers in a supervised manner. A competitive layer will automatically learn to classify input vectors. However, the classes that the competitive layer finds are dependent only on the distance between input vectors. If two input vectors are very similar, the competitive layer probably will put them into the same class. There is no mechanism in a strictly competitive layer design to say whether or not any two input vectors are in the same class or different classes. LVQ networks, on the other hand, learn to classify input vectors into target classes chosen by the user. The architecture of the network can be found in Appendix F.

An LVQ network has a first competitive layer and a second linear layer. The competitive layer learns to classify input vectors in much the same way as the competitive layers. The linear layer transforms the competitive layer's classes into target classifications defined

University of Malaya

by the user. We will refer to the classes learned by the competitive layer as subclasses and the classes of the linear layer as target classes.

The competitive transfer function accepts a net input vector for a layer and returns neuron outputs of 0 for all neurons except for the winner, the neuron associated with the most positive element of net input n 1. The winner's output is 1. If all biases are 0 then the neuron whose weight vector is closest to the input vector has the least negative net input, and therefore wins the competition to output a 1.

LVQ learning Rules

LVQ learning in the competitive layer is based on a set of input/target pairs as below:

$$\{p_1, t_1\}, \{p_2, t_2\}, \ldots, \{p_Q, t_Q\}$$

Each target vector has a single 1. The rest of its elements are 0. The 1 tells the proper classification of the associated input. For instance, consider the following training pair.

$$\left\{p_1 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}, t_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}\right\}$$

Here we have input vectors of three elements, and each input vector is to be assigned to one of four classes. The network is to be trained so that it will classify the input vector shown above into the third of four classes.

To train the network an input vector p is presented, and the distance from p to each row of the input weight matrix IW {1,1} is computed. The hidden neurons of layer 1 compete.

University of Malaya

Suppose that the ith element of n 1 is most positive, and neuron i* wins the competition. Then the competitive transfer function produces a 1 as the i*th element of a 1. All other elements of a 1 will be 0.

When a 1 is multiplied by the layer 2 weights LW {2,1}, the single 1 in a 1 selects the class, k* associated with the input. Thus, the network has assigned the input vector p to class k*. Of course, this assignment may be a good one or a bad one. We will adjust the i*th row of input weights IW {1,1} in such a way as to move this row closer to the input vector p if the assignment is correct, and to move the row away from p if the assignment is incorrect. So if p is classified correctly, the new value of the i*th row of IW {1,1} are computed using the equation below:

$$ {}_{i^*}\mathbf{IW}^{1,1}(q) = {}_{i^*}\mathbf{IW}^{1,1}(q-1) + \alpha(\mathbf{p}(q) - {}_{i^*}\mathbf{IW}^{1,1}(q-1)). $$

On the other hand, if p is classified incorrectly,

$$ (a^2_{k^*} = 1 \neq t_{k^*} = 0). $$

the new value of the i*th row of IW {1,1} are computed using the equation below:

$$ {}_{i^*}\mathbf{IW}^{1,1}(q) = {}_{i^*}\mathbf{IW}^{1,1}(q-1) - \alpha(\mathbf{p}(q) - {}_{i^*}\mathbf{IW}^{1,1}(q-1)) $$

Such corrections move the hidden neuron towards vectors that fall into the class for which it forms a subclass, and away from vectors, which fall into other classes. [17]

University of Malaya

2.3 Intelligent Agent

An agent is a software entity which has some degree of autonomy, carries out operations on the behalf of a user or another program, and in this process, represents or has knowledge of the user's goals and wishes [6].

2.3.1 What Is An Agent?

The intelligent agent is a concept that has been around for approximately 25 years. The definitions of an agent fall everywhere along a continuum; from simple macro in which the user enters a few parameters to truly intelligent agents, which demonstrate learning ability and artificial intelligence. Currently, most intelligent agents fall somewhere between the former description and the middle ground of the continuum. In general, though, all agents seem to have some similar characteristics.

An example of an agent task would be if the user wished to find the lowest price for a specific item, such as music CD. Rather than the user having to explore the Internet for endless hours trying to locate the least expensive price, the agent would be free to do as he wished. Monitoring the search would not be necessary and a list of low prices with store identification would be returned.

The qualities of an agent can be broken down further into three main dimensions: agency, intelligence, and mobility. Agency describes the degree of independence, which an agent exhibits. At the least, the agent must be able to operate on the Internet while the user is disconnected or not in the process of interacting with the World Wide Web. The most

University of Malaya

basic agent represents the user in some way and proceeds to complete a task with that information in mind. More complex agents can interact with other programs and data. Eventually, as the field of artificial intelligence continues to grow, this independent agent will be able to cooperate with others agent in a mutually beneficial way.

Intelligence is the amount of learned behavior and possible reasoning capacity that an agent may possess. At the most basic level, the agent may follow a set of rules that are pre-defined by the user. The agent can then apply these rules to the Internet at large. An agent who followed a model would demonstrate the next level of intelligence. A user model would not have to be as specific as a set of rules and would include a plan to retrieve the information as well. Finally, the most intelligent agents will be able to learn. They will also be able to adapt to their environment, in terms of user requests and the resources available to the agent. These agents may even make their own relationships with other agents and create new agents with higher reasoning capabilities. All of this interaction will benefit the user by providing better information; yet will be invisible to the user. Basically, their needs will be met and the agent will have taken care of the details.

Mobility is the dimension, which makes agents useful in Internet applications, yet poses many privacy and security issues. An agent can be static and reside on the client computer to manage and gather information on that one machine. A more mobile agent may be composed on one computer and then sent to another to gather data. It then returns the data to the client and is sent out to gather another piece of information. This pattern

repeats until the agent has exhausted the search parameters or the user stops it. This arrangement has the advantage of good security.

The other option related to mobility is an agent which travels from computer to computer gathering information, but does not return to the client computer until it has collected all of the requested data. In other words, this agent accumulates and carries the data in the process of its search. While security is not as good on this form of agent, it will have the distinct advantage of allowing agents, which have similar information and goals to interact. There may even be places where the agents will "meet" to share information and collaborate on their individual quest for data.

2.3.2 The Difference Between Search Engine and Intelligent Agent

At first glance, search engines such as Yahoo, Lycos, and WebCrawler seem to match some of the capabilities of basic agents. The key difference though is that an agent is more interactive and can accomplish many tasks at many different locations. For example, if searching on a keyword using a search engine such as Yahoo, user will get a list of matches. User can then follow those links and possibly get their information. If user were to use an agent, on the other hand, the agent could submit the user keyword(s) to many search engines, follow the corresponding links, and gather the information, all without any intervention from the user.

University of Malaya

2.3.3 Shopping Agent

An agent can be applied in many domains such as in industrial application, medical application, commercial, application and so on. Ever since the growth of e-commerce, shopping online has become a popular way to buy things. Consumers are confronted with information overload when they go shopping on the Internet. Consumers want to be able to find a broad selection when looking for a product; however, the selection must be edited down to reasonable and manageable proportions. This editing feature is the job of the intelligent agent. Without this editing feature, buying a simple music CD would be quite an experience.

BargainFinder is an intelligent agent which shops for the lowest price of a CD. Although BargainFinder has made a valiant effort, there have been problems. The speed is not superior, and it lacks the ability to retain knowledge. Another problem is that many companies do not want to be judged on price alone, causing them to block BargainFinder from their site [9].

Firefly also helps the user find music, but it is more sophisticated in that it learns the user's musical preferences. Firefly asks users to rate a number of different music artists, then it suggests other types of music that the user might like. The suggestion is based on correlations with what other people say they enjoy listening to, versus using artificial intelligence. Firefly could be used in conjunction with BargainFinder, or even better, combined. If this was implemented, user could have the agent suggest a title and then find the best deal on the CD [10].

2.4 Electronic Commerce and The Internet

Internet is a wide area network connecting thousands of disparate networks in industry, education, government and research. The Internet network uses TCP/IP as the standard for transmitting information [8].

The Internet began with the development of ARPANET and subsequent support by the Defense Advanced Research Projects Agency (DARPA) for the development of additional networks to support military users and government contractors. At the beginning, Internet is used for communication through e-mail between the employees in a corporation. However many companies now are starting to do business over the Internet because it provides a greater opportunity for business expansion and more simplified transactions. Thus, electronic commerce is a new chapter in the context of Internet.

Electronic commerce is defined as any form of business or administrative transaction or information exchange that is executed using any information and communications technology (ICT). Following this definition, information on these pages includes the latest technologies for enabling trade over the internet, the more established practices of conventional EDI (electronic data interchange) and bar coding, as well as the business reasons for implementation of these technologies [7].

Electronic commerce represents a market worth potentially hundreds of billions of dollars in just a few years and provides a tremendous opportunity for businesses in various areas. It connects business process between various entities in a company supply chain,

resulting in reduced processing costs and enabling appropriate access to business partner systems.

Today, electronic commerce is evolving into a more cost-effective and powerful way to do business whereby goods and services are exchanged for value on the Internet. This new paradigm of doing business has enabled companies to market and sell their products all around the world in 24-hours a day.

Recent significant progress in Artificial Intelligence for e-commerce includes:

1. Practical shopping agents, including Web portal services, which use knowledge representation, decision analysis, machine learning, and information retrieval techniques.
2. Practical recommending services, e.g., e-storefronts that use collaborative filtering.
3. Practical data mining by sellers to learn customers buying patterns.
4. Practical customer-service help, including agent techniques to categorize and route e-mail, do case-based associative retrieval and make suggestions.
5. Theory of economic decision-making markets, negotiations, and contracts, including from the viewpoints of resource-bounded intelligence, game theory, distributed AI, negotiation, probabilistic and uncertain reasoning, and decision analysis.
6. The theory and practice of auctions.
7. Agent communication languages, including negotiation languages and protocols and knowledge interchange and the use of XML-encoded domain ontology and communication languages.

University of Malaya

8. Web information retrieval and information integration, including using NLP, text analysis and machine learning.
9. Online product/service catalogs, e.g., techniques to aggregate catalogs.

University of Malaya

2.5 Software Used

This project will used the following software:

- Matlab Released 11 (Neural Network and Database Toolbox)
- Active Server Pages (ASP)
- JavaScript
- VB Script
- Microsoft Access 2000
- Hyper Text Markup Language (HTML)

The reasons and purpose of the use of this software will be discussed in Chapter 3 Methodology.

The next chapter is about the methodology used for developing this project.

University of Malaya

CHAPTER III

METHODOLOGY

This chapter will discuss about the user and system requirement and the waterfall model. The hardware and software requirement will also be covered here.

3.1 System And User Requirement

The system requirement needs to be drawn out to provide a guideline when developing a system. A requirement is a feature of the system or a description of something the system is capable of doing in order to fulfill the system purpose [11]. There are two types of requirement, which is as follow:

- Functional requirement
- Non-functional requirement

3.1.1 *Functional Requirement*

A functional requirement describes an interaction between the system and its environment. It also describes how the system should behave when given a certain stimuli. The functional requirement for this system are stated below:

University of Malaya

General User Section

This section is responsible to communicate with user in getting information from user to proceed the searching and price prediction, and displays the outcome to the user.

i. *Display Module*

This is a front-end design, which is responsible for the interactions between user and the system. It consists of the following form:

Search Form

The purpose of the search form is to enable the user to input their required criteria such as state, location, area, house type and number of bedroom of a house. These criteria will be combined to enable a Boolean search in the search module.

Results Page

This page includes a listing of entire search results, which is the relevant information of the houses and its predicted house price in the market.

Sending feedback form

User can give comment on the system by using the feedback form provided. The collected comments will be used to improve the functionality of the system.

University of Malaya

ii. *Agent Section*

There are two parts in the agent section, which play an important role in the back-end of the system.

Filtering Module

This module aims to ensure that only relevant information is return to the user. The filtering module is responsible to filter the data in the database and comes out with the user request.

Prediction Module

The purpose of this module is to assist the user in decision-making by predicting the value of the house in the current market where user's query terms are used.

iii. Administration Section

This is an offline section that plays an important role in system enhancement.

Training Module

This module aims to train the filtering agent and prediction agent in producing the best results.

University of Malaya

Feedback administrator

The administrator is responsible on collecting user feedback, which may be useful for system improvement and enhancement in the future.

3.1.2 Non-functional Requirement

A non-functional requirement or constraints describes a restriction on the system that limits choices for constructing a solution to the problem. These constraints usually narrow the selection of language, platform or implementation technique and tool. Below are the non-functional requirements of the system:

i. *Maintainability*

Maintainability is the degree to which the system can be cost-effectively made to perform its functions in a possibly changing operating environment. The system are easy to modify and test in updating process to meet the new request, correcting errors, or move to a different computer system.

ii. *Reliability*

The degree in which the system operates in a user-acceptable manner when used in the environment for which it was designed, which does not produce dangerous or costly failure when it is applied in a reasonable manner.

University of Malaya

iii. Efficiency

Implementation of the system corresponds to the most cost-effective computing resource utilization, where process that can be called or accessed in an unlimited number of times to produce similar outcomes at a creditable pace or speed.

iv. User friendliness

The design of the system and its interface should be user friendly and easy understanding by all level of the Internet users. Generally, the design of all the interfaces should conform to the following criterions:

- Consistent, in terms of screen design and error messages displayed.
- High degree of understandability and avoid memorization of events and commands.

v. Simplicity

Forms and screens are kept properly uncluttered in a manner that focuses the user attention.

vi. *Understandability*

Coding method used, allow other programmer to understand the logic of the program flow.

University of Malaya

3.2 Waterfall Model With Prototyping

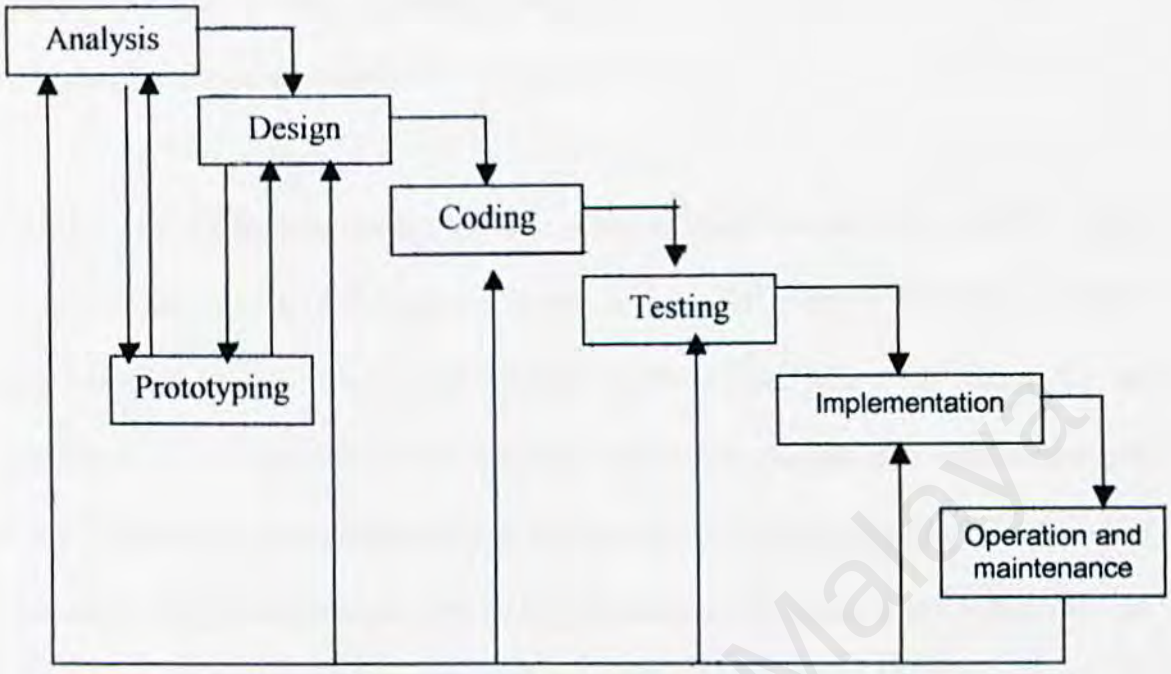


Figure 3.1 Waterfall Model With Prototyping

Conventional development and prototyping may be combined. The Waterfall model with prototyping is chosen because the strengths of each can be achieved on a single project. This model is actually the classic waterfall model combined with the prototyping approach in its early stages as shown in figure 3.1 above.

In the proposed development strategy, the waterfall model will serve as the base for the development because the steps of this model are very similar to the generic steps of software development process that are applicable to all software engineering paradigms. It also provides a template into which methods for analysis, design, coding, testing and maintenance can be placed. Prototyping will be involved in the early stages of the waterfall model where there is a need for experimentation and learning before

commitment of any resources to develop the full-scale system. Prototyping will not be involved in the later stages of the development because its major drawback in increasing the opportunities to produce negative effects on structural factors such as performance, design quality and maintainability if not carry out properly.

In the case of the proposed project, an idea solution might be one that combined rapid results (from prototyping) with stability (from the waterfall model). The stability of the classic waterfall model is very much needed in this project. The main reason for the incorporation of prototyping into the waterfall model is to rapidly elicit and experiment with user interface requirements and usability factors. Prototyping approach is also ideal in the sense that the developer has neither complete information/understanding nor experience in developing this type of system.

The waterfall model with prototyping approach that will be adapted in the proposed project encompasses the activities at system analysis, system design, coding, testing and implementation. Each activity is discussed below:

Analysis

The goal to the system analysis is to understand the proposed system and to establish system requirements. The system analysis phase is concerned with data gathering and data analysis.

Data will be gathered from sources like written materials, Internet as well as observation and examination of others intelligent agent systems. The iterative process of prototyping-revision will be Data Flow Diagram (DFD) is chosen to analyze the collected data because it enables the information domain and functional domain to be modeled at the same time. It is to be used to graphically show the flow of the data through the system.

The most important outcome from this phase will be an accurate system requirement specification.

Design

The system design phase is the phase in which requirements produced in the previous phase are translated into a representation of the system. This phase will be concerned with user interface design, database design and system design.

The interactive process of prototyping-revision will be used to revise the design of the user interface. Interface prototypes will be built using Microsoft Visual Interdev. Entity-relational (E-R) modeling will be involved in the logical design of the Microsoft Access database. In system design, structure chart will be involved in structuring the system's modules and flow chart might be used to depict the design of procedural details.

University of Malaya

Coding

This stage translates and implements the detail design representation of the system into programming realization. Scripting languages such as VBScript and HTML might be used in coding the information and functional domain as well as the control of the proposed system. Microsoft FrontPage is the proposed web-authoring tool that will be used to create web pages while Microsoft Access 97 will be used to develop the database of the system.

Testing

Testing will be a critical step in assuring the quality of the developed system and will represent the ultimate review of specification, design and coding. First, unit testing will be performed to verify each program module. Next, integration testing is performed. It is to integrate unit-tested program modules and conduct tests that uncover errors associated with the interfacing of those modules. Validation test succeeds when the system functions in the manner that is reasonably expected.

Implementation

The finally stage of the development will be system implementation. The system will be implemented on its target software and hardware requirement. The whole system will be revise to uncover the necessity to add further enhancements.

University of Malaya

Operation and Maintenance

Maintenance process should be an ongoing activity in real development. Monitoring a necessary adjustment continue so that the system produces the expected results. However, system enhancements and maintenance will only be carried out in the proposed project if time constraint allowed.

University of Malaya

3.3 Development Environment

The development environment of a system is the software and hardware requirement of the system.

3.3.1 Hardware Requirement

The system needs the following hardware requirement for its operation.

- 96MB of RAM
- 2.1GB of Hard Disk
- VGA color monitor
- Keyboard and mouse

3.3.2 Software Requirement

In order to develop this system, a few software products will be used. These software products are divided into a few categories, which are software for Operating System, Web Server and Web application programming. The usage and advantage of each of this software are stated below:

3.3.2.1 Operating System

Microsoft Windows NT Server 4.0

The Microsoft Windows NT Server 4.0 is one of the powerful operating system for business computing nowadays. There are several reasons why Windows NT Server 4.0 is used: -

University of Malaya

- It provides an interface, which is easy to use and this helps user to work easier and faster.
- Windows NT Server 4.0 is easy to manage and control as it includes remote management and troubleshooting tools as well as allowing administrator to implement policies and standards for system desktop configuration.
- Its productivity and compatibility are very suitable for this system because Windows NT Server 4.0 ensures high performance for 32-bit programming. All win-16 Windows based programming have the preemptive multitasking capabilities of Windows NT and can be run in a separate address space for better responsiveness and reliability.
- It meets the reliability standards to run critical line of business programming. It also allocates separate memory space for 16-bits applications, so if one 16-bit application fails, it will not bring down the other applications. Furthermore, it will also protect critical Operating System code, device drivers and data from application.
- Windows NT supports workgroup and networking. It has the built-in file sharing and print-sharing capabilities for workgroup computing. This Operating System also has open network system interface that is compatible with Netware, Novell, UNIX and LAN manager as well as Microsoft Windows for Workgroup, Windows 95 and standard x.86 environment. Windows NT computers can made up to 10 simultaneous connections for sharing files and printers.

University of Malaya

- It allows Object Linking and Embedding (OLE). In other words, it can combines information from several applications into one compound document using the special ODE capabilities of win-based application.
- It has built-in tools for internetworking and intranet working like TCP/IP, Microsoft Internet Explorer, Microsoft Peer Web Services and so on.
- Furthermore, it also enables the capabilities of integrating application on a single computer or even across multiple computer using COM and DCOM.

3.3.2.2 Web Server

Internet Information Server 4.0 (IIS)

Internet Information Server 4.0 is a high performance web application server for Windows NT Server 4.0. It brings many advanced capabilities to web professionals, both as a web server for corporate intranets and public Internet sites as well as a platform for the next generation of line-of-business applications.

This web server also incorporates World Wide Web (WWW), File Transfer Protocol (FTP), Microsoft Index Server and Secure Socket Layer (SSL) services. Furthermore, it also provides comprehensive set of tools for managing the web server and its components. [13]

With IIS, it is an easy task to build scalable and reliable application for the web. Developer can easily integrate the benefits of transactions into web applications using IIS. IIS also brings together the advantages of Windows NT Server, with a range of

University of Malaya

services for client/server development and the standards of the Internet to create a true web platform for distributed application.

3.3.2.3 Neural Network Development Tools

Matlab

Matlab is used to develop the neural network of the system. This software is chosen because it has toolbox that provides all the libraries to develop a neural network, thus it is not necessary to create a neural network from scratch.

Neural Network Toolbox

Here, two toolboxes will be used, which are the Neural Network Toolbox and the Database Toolbox. The Neural Network Toolbox is used to implement the neural network as the back-end of the system. This toolbox provides many kinds of networks such as Perceptron, Self-Organizing Network, Adaptive Linear Filters, Learning Vector Quantization (LVQ), feed forward network and so on. It also supports custom network, where users can create their own network type from scratch and design the network's architecture. This means that, users can define the learning functions which is available in the toolbox and number of inputs, outputs, bias, transfer functions and so on. Thus, users can define the attributes of the network themselves. Meanwhile, if the available types of network are used, the network architecture is already set and users only have to set the training parameters to use the network.

University of Malaya

For this system, the LVQ network will be used because it is used to performed classification. The architecture of the LVQ network is already defined in the Neural Network Toolbox and users only have to set the training parameters. To create the LVQ network, users just have to call out this function: *'newlvq'* and set the necessary arguments for it. After that, if the users want to set the training parameters, they can just call the network attributes such as *net.trainParam.epochs* = 3000 to set the number of epochs where *net* is the network object, *trainParam* is the function to adjust training parameters and *epochs* is the attributes to set the particular number of epochs specify by users. All of these features enable easy manipulation of the creation of the neural network.

Database Toolbox

Other than the Neural Network Toolbox, the Database Toolbox will also be used to develop the neural network. This toolbox can enable data to be imported and exported from the Matlab environment to other environment. It is used here because the interface of the system is developed using other software like HTML, VBScript and so on, which are not compatible with Matlab environment. Thus, the Database Toolbox is used to connect to the database where all the data are stored there for retrieval and updating.

The data that is produced by the neural network can be exported to the database where other modules can retrieve from the database and display the data to users. On the other hand, data that was input by the user can be imported to the Matlab environment for the use of the neural network can also be accessed through the Database Toolbox. The

University of Malaya

utilization of the Database Toolbox, enables the flow of data to or from the Matlab environment to other environment and reduces a lot of technical work. It is also chosen, because this system uses Microsoft Access 2000 as its Database Management System, which is supported by the Database Toolbox.

3.3.2.4 Web Application Programming Languages

VBScript

VBScript is an interpreted script language from Microsoft that is a subset of its Visual Basic programming language. In general, script languages are easier and faster to code in than the more structured, compiled languages such as C and C++ and are ideal for smaller programs of limited capability or that can reuse and tie together existing compiled programs. The function of VBScript is to make web application more dynamic. VBScript is chosen because it is more reliable to work with ASP.

It is implemented as a fast, portable, lightweight interpreter for use in WWW browsers and other application that use Microsoft ActiveX controls, Automation Servers and Java applets. VBScript currently is available as a part of Internet Explorer and IIS. [14]

When used with Internet Explorer, VBScript is directly comparable to JavaScript. Like JavaScript, VBScript is a pure interpreter that processes source code embedded directly in the HTML. VBScript code, like JavaScript does not produces standalone applets but is used to add intelligence and interactivity to HTML document. For programmers who

University of Malaya

already know Microsoft Visual Basic, VBScript is a valuable alternative to JavaScript in developing web pages.

JavaScript

JavaScript is an interpreted programming or script language from Netscape. It is used in Web site development to do such things as: Automatically change a formatted date on a Web page, cause a linked-to page to appear in a popup window, cause text or a graphic image to change during a mouse rollover. JavaScript is useful for adding interactivity to the World Wide Web because scripts can be embedded in HTML files. JavaScript can also be run at the server as in Microsoft's Active Server Pages before the page is sent to the requestor.

Active Server Pages (ASP)

ASP is a script that runs in Microsoft Web Server. It is used to generate HTML scripts for the client browser. Compared to CGI, ASP is easier to use and is more flexible in changing code as no compilation is involved and so it is chosen as the main development tool for server run script.

Active Server Pages (ASP) is the latest server-based technology from Microsoft for building dynamic and interactive web pages. The basic of ASP is the Microsoft IIS software.

ASP is not an application. It is a VBScript and JavaScript interpreter that is integrated with IIS, together with an interface for other custom component. It is also able to include

other web pages components like ActiveX controls and Java Applets. Therefore, ASP is considered as a glue technology, which binds together other various server-based system to help build interactive web pages. [16]

ASP is used to build web pages because:

- It is suitable for publishing and collecting data on the web.
- It provides way for building secure transactions, server-based applications and web sites.
- ASP works together with Windows NT and IIS to provide a comprehensive set of key software technology, which enable secure exchange of information over public network, access control to server resources and confident identification of server and client.
- Furthermore, ASP also provides ADO, one of ASP component that allows easy but powerful connections to be made to almost any database system for which an ODBC driver is available.
- It has pre-build ASP components, which provide plug-in object that perform specific tasks.
- Besides this, ASP can interact with almost any existing dynamic web page technology such as CGI, ISAPI and scripts written in PERL, Python and Awk.
- It also suitable for building multi-tier Internet and Intranet applications.
- Supports client-server programming. Furthermore, the combination of ASP, client-side scripting and object can be used to create client-server applications.
- With ASP, it is able to create client-side code dynamically on server.

University of Malaya

ASP Compare to CGI Applications

ASP provides all the functionality of CGI applications in an easier-to-use and more robust environment. ASP is an easier way to access information in a form not readable by the client (such as an SQL database) and then acts as a gateway between the two to provide information that the client can view and use.

With CGI, the server creates as many processes as the number of client requests received. The more concurrent requests there are, the more concurrent processes created by the server. However, creating a process for every request is time consuming and requires large amount of server RAM.

In addition, this can restrict the resources available for sharing for the server application itself, slowing down performance and increasing wait time on the web. ASP instead, runs in the same process as the web server, more handling client requests faster and more efficiently. It is much easier to develop dynamic content and web applications with ASP.

ASP Compare to ISAPI Application

ISAPI application requires all of the programming and layout to be contained in a DLL file, written in C++. Therefore, it is more difficult to create and maintain with ASP files and HTML writer can script an external component and format the output. ASP separates the layout and design from the business logic.

University of Malaya

ASP Compare to PERL

PERL and other scripting languages are not robust development tools by themselves. ASP provides a familiar framework and objects for building complex application that requires data from relational database and legacy services. ASP supports virtually any scripting language to build these applications. Third parties are currently developing additional scripting engines, which will be released when it is ready.

Hypertext Markup Language (HTML)

To publish information for global distribution, one needs a universally understandable language, just like a kind of publishing mother tongue that all computers may potentially understand. The publishing language used by the WWW is HTML.

It is a basic tool for that is needed for the development of client browser run script. HTML is the set of "markup" symbols or codes inserted in a file intended for display on a World Wide Web browser. The markup tells the Web browser how to display a Web page's words and images for the user. The individual markup codes are referred to as elements.

HTML gives author the means to do the following:

- Publish online document with headings, text, tables, lists, photos and so on.
- Retrieve online information via hypertext links at the click of a button.
- Design form for conducting transactions with remote services, for use in searching for information and so on.

University of Malaya

- HTML also includes spreadsheets, video clips, sound clips and other applications directly in its documents.

Most people agree that HTML document should work well across different browsers and platforms. Achieving interoperability lowers costs to content providers since they must develop only one version of a document. If the effort is not made, there is much greater risk that the web will be devolved into a proprietary world of incompatible formats ultimately reducing the web's commercial potential for all participants.

Each version of HTML has attempted to reflect greater consensus among industry players so that the investment made by content providers will not be wasted and that their document will not become unreadable in a short period of time. HTML has been developed with the vision that all manner devices should be able to use information on the web. PC with graphics displays of varying solutions and color depths, cellular telephone, hand held devices, devices for speech for output and input, computers WITH high or low bandwidth and so on. [15]

3.3.2.5 Web Application Development Tools

Microsoft Visual Interdev 2000

Microsoft Visual Interdev 2000 is the newest member of the visual tool family, is an integrated development tool for building dynamic web applications accessible by any web browser on any platform. It includes an integrated development environment,

University of Malaya

database connectivity tools, programmable components, site management and publishing capabilities, a personal web server, content creation tools and many more.

Visual Interdev also includes a variety of development features for integrating client-server and web technology. These features are enabled through Visual Interdev support for ActiveX controls, seamless database connectivity to any ODBC data source, support for building and testing large system and comprehensive support for the development of web application.

Furthermore, it this tools also provides a rapid visual development environment for building ASP. It can easily integrate ActiveX server components written in Visual J++, Visual Basic, Visual FoxPro and Visual C++. Using Visual Interdev with ActiveX server components, a developer can easily creates multi-tier web applications. ActiveX server component provides a convenient and effective way to tightly integrate a web application with existing Internet system.

Visual Interdev delivers a comprehensive set of tightly integrated database tools for web developers. The database connectivity features are based on the industry standard ODBC including Oracle, Microsoft SQL Server, Microsoft Access, Microsoft Visual FoxPro, Informix, Sybase, IBM DB/2 and so on. In addition, with Visual Interdev, a developer can create a scalable database solution because it leverages ASP. The core database components of Visual Interdev include ADO, Integrated Data View, Design-Time ActiveX Controls, Database Wizards, Query Designer and so on.

University of Malaya

A Visual Interdev project consists of a live web site. When developer opens a project, they are actually opening a live view of a web site, as it exists on the web server. The IDE is thus a complete web site management tool that allows the developer to easily modify the structure of a web site and to edit, add, move, rename and delete files and folders on the web sites. It can also open multiple web sites at the same time.

3.3.2.6 Database Management System

Microsoft Access 2000

It is used to serve as Database Management System for the system. Access is chosen because it can provide relational database power to manipulate information. It also supports SQL statements and can be easily integrated with programming languages used to develop the system. Microsoft Access is used because it is more suitable for regular database compare to Oracle, which is more suitable for building a large database. Furthermore, Matlab can support Microsoft Access using its Database Toolbox. This is an important feature because data can be imports and exports to Matlab environment from or to Access database.

The following chapter will explain about the design of the system. The modules of the system will be presented in that chapter.

University of Malaya

CHAPTER IV

SYSTEM DESIGN

In this chapter, the implementation step of the system will be explained and overview design of the whole system will be presented first. Then the display module, search module and filtering module will be discussed further.

4.1 Implementation Step

The Intelligent Agent System was designed to helps user to obtain the required house information from the World Wide Web and propose a predicted price for the specified house to assists user in decision-making. The implementation steps includes the following chain of events:

- Solicit a subject from user query,
- Sent the user inputs and the houses information stored in database to the network,
- Filter out irrelevant Web pages,
- Display the relevant result to users,
- Predict the price for the query subject,
- Display the predicted price to user.

Once the user fill in and submit the search form, the house criteria will be uploaded to the Search Agent. This in turn, will send the user inputs and the houses information stored in database to the neural network, which is the Filtering Agent. The Filtering Agent, with

University of Malaya

embed neural network, given a set of correct weight, will attempted to map the query terms (user input), which is represent as the input vectors to the output vector, which could indicate the "relevance" of the input vector to a certain house required.

The relevant Web pages will be select by the Filtering Agent as the output. On the other hand, the Prediction Agent, another neural network, which will predict the value of house required by user. In this case, the house price is the target output, and criteria such as location, house type, square feet, number of bedroom, and land size are the variables used to predict the selling price. After this network is trained with a set of data, it could predict a selling price for a specific house. This predicted price is used to assist user in decision-making. Finally, the relevant information that was filtered out previously and the predicted price will be displayed to user.

University of Malaya

4.2 System Overview

This system is divided into four major modules that are as follow:

- Display module
- Filtering module
- Prediction module
- Training module

Display Module

This is a user interface that consists of two main screens:

i. *User Input Screen*

Search form to get query from user to help them to obtain the required information.

ii. *Output Screen*

This page is used to display the information of the houses from the search result. The predicted current market price is also presented to assist user in decision-making.

Filtering Module

This module will use neural network to do the filtering of the data in the database. It will filter out the relevant information, which is required by user from the database and display it to users.

University of Malaya

Prediction Module

The function of this module is to predict the current market price of the house that is required by the user. The price will be determined by using neural network with an input set of criteria input by the user in the input screen.

Training Module

It is an offline module that aims to train the neural network in the prediction module. This training is used to train the network in order to make it more accurate by frequent training.

University of Malaya

The system data flow diagram is shown below:

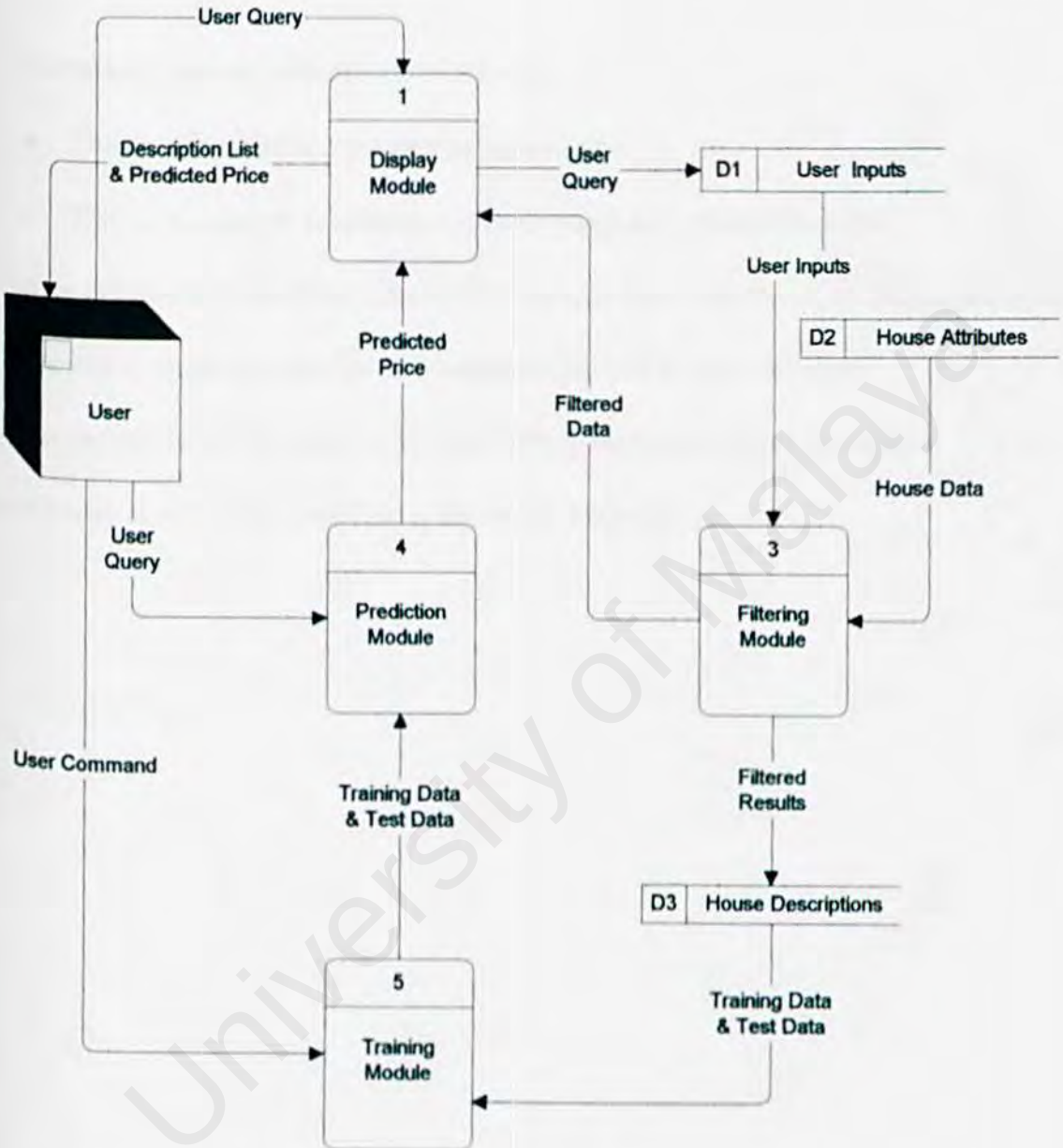


Figure 4.1 Data Flow Diagram Of The System.

University of Malaya

4.3 Display Module

The display module has the following features:

i. Contained two main interfaces which are:
   - The user input screen for user to input query
   - The output screen to display searched result and predicted prices.

ii. Also contain links to other related sites such as bank, insurance, architect and so on.

iii. Provide a feedback form for user and also has site to train the agent.

iv. An on-line help will also be provided to help the user in using the system.

The example of user input interface is shown in Appendix A.

University of Malaya

4.4 Filtering Module

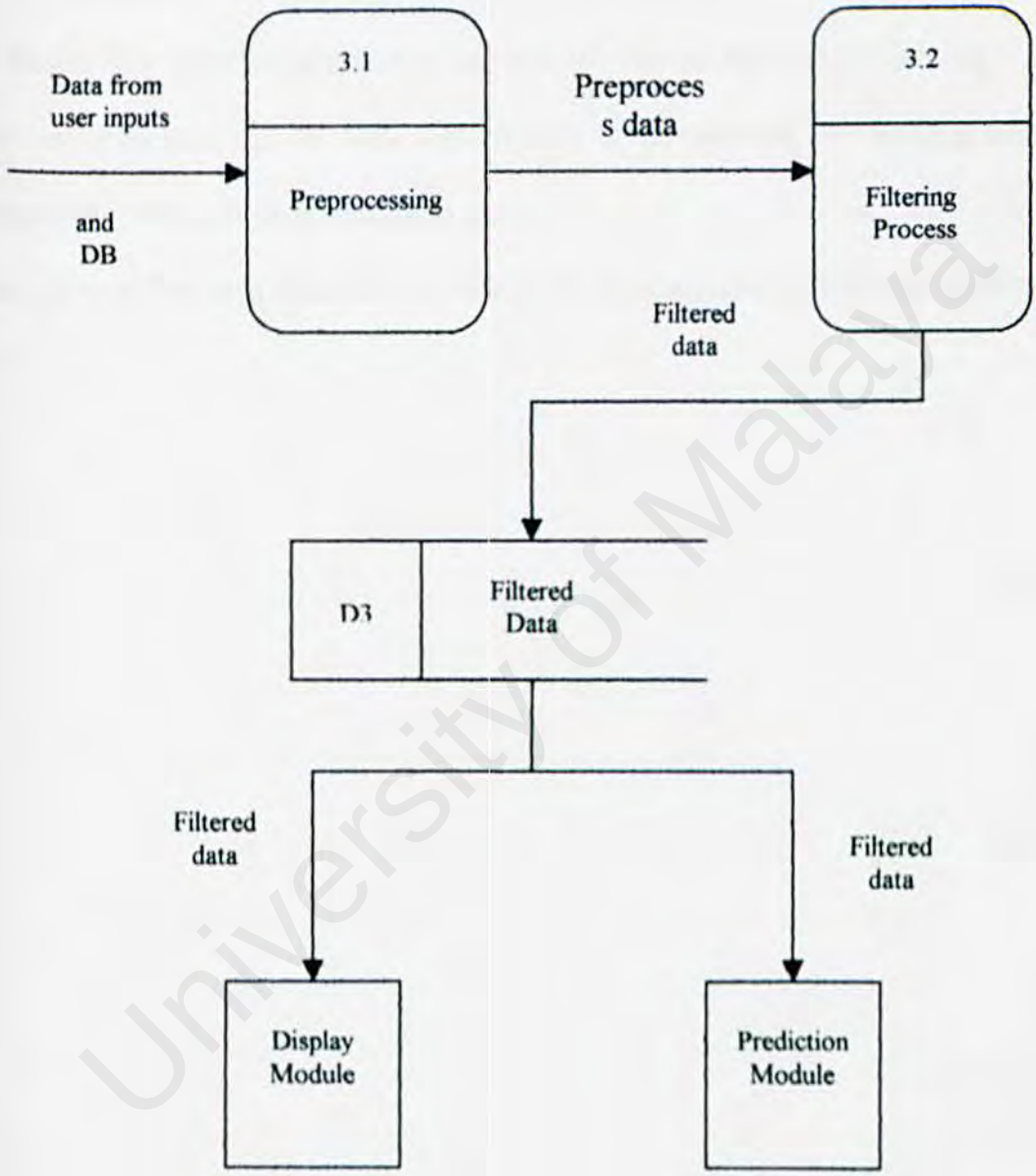


Figure 4.2 Data Flow Diagram of The Filtering Module

University of Malaya

This module is used to get relevant information from the web pages returned by the World Wide Web. The procedure of the filtering module is as follow:

1. Get the user inputs and the data of houses store in database.
2. The data will be preprocessed first before it is fed into the network for filtering.
3. After the preprocessing, the data will be sent to the network for filtering out the relevant data, which is required by the users.
4. These relevant data will then be sent back to the database and will be displayed to the user.

University of Malaya

4.5 Neural Network Design

The LVQ network has been chosen as the neural network to be implemented in the Filtering Module. LVQ is a method for training competitive layers in a supervised manner. A competitive layer will automatically learn to classify input vectors. However, the classes that the competitive layer finds are dependent only on the distance between input vectors. If two input vectors are very similar, the competitive layer probably will put them into the same class. There is no mechanism in a strictly competitive layer design to say whether or not any two input vectors are in the same class or different classes. LVQ networks, on the other hand, learn to classify input vectors into target classes chosen by the user.

There are several reasons why the LVQ network type is chosen as the neural network of the system:

- The LVQ network is used to perform classification. Although, Perceptron can also be used to do classification, the LVQ network can do non-linear problem classification, which the Perceptron cannot do.
- This network is already provided in the Matlab Neural Network Toolbox. Thus, it will be more convenient to just call the function to create the network. The toolbox also provides the training parameters, which can be used just by calling the function out. Furthermore, the LVQ learning rules is also embedded in the toolbox and users do not have to implement the learning rules, which can be quite complex to the network.

University of Malaya

- The neural network is used here to determine whether the information is relevant to the user or not. Thus, the information that is available has to be filtered in order to determine which one is relevant to users. The LVQ network can be used to do classification, where the information is classified as relevant or not relevant to users.
- It can be used to do classification according to users selection of target classes, which is suitable for the filtering module because the relevance of the information is determined by the a few criteria. This kind of classification cannot be done in if the Self-Organizing network (SOM) or the other unsupervised learning networks are used because these network perform classification automatically without any training pattern.

University of Malaya

LVQ Network Architecture Design

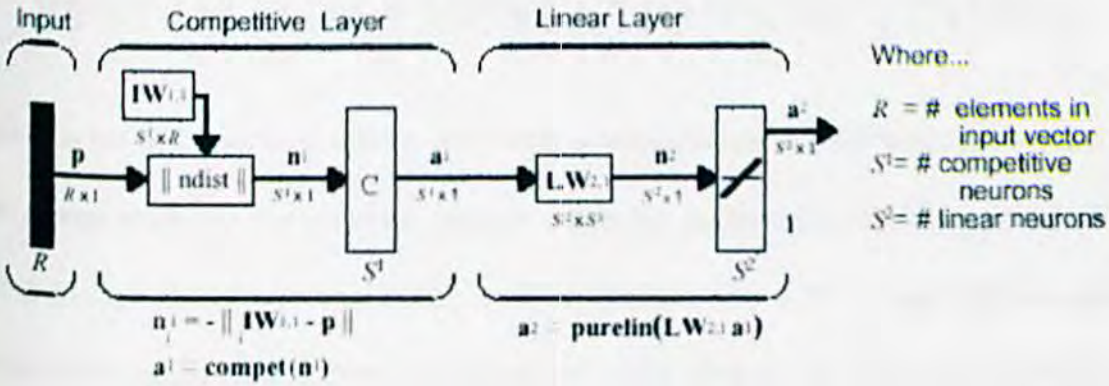


Figure 4.3 LVQ Network Architecture Design

IW {1,1} = Input Weights.

LW {2,1} = Layer Weights.

C = Competitive Transfer Function (*compet*).

a (1) = Output from competitive layer.

a (2) = Output from linear layer.

n (1) = Input to *compet* transfer function.

n (2) = Input to *purelin* transfer function.

The next chapter is the conclusion of the project where the summary of the project and the expected outcome of the system will be discussed.

University of Malaya

CHAPTER V

SYSTEM IMPLEMENTATION AND TESTING

This chapter will discuss about the steps and methods taken to implement the system that was design earlier in the previous chapter. After the implementation, the system will be tested to look for errors. The testing will be divided into a few types such as module testing, integration testing, function testing and other testing. All of this type of test will be discussed later.

5.1 Developing Environment

The developing environment for a system is the tools used to develop the system. There are two kinds of tools used, which are the hardware tools and the software tools. Both tools are described below for the server side and the client side.

5.1.1 Hardware Tools

The following list is the hardware tools used to develop the system:

Server	Client
1. Pentium III Processor.	1. Pentium III Processor.
2. 96 MB RAM	2. 96 MB RAM
3. 2.1 GB Hard Disk Drive	3. 2.1 GB Hard Disk Drive
4. Ultra VGA 1024 Monitor	4. Ultra VGA 1024 Monitor
5. Mouse	5. Mouse
6. Keyboard	6. Keyboard

University of Malaya

7. Local Area Network	7. Modem

Table 5.1 Hardware Requirements Tools

This configuration is sufficient enough to run the system.

5.1.2 Software Tools

The software that has to be implemented in the server and the client side to run the system is stated below:

Server Machine

1. **Windows 98/NT**

   This is the operating system or platform used in the server machine to run the system.

2. **Matlab Released 11 or higher**

   The Matlab must be installed in the server machine in order for the back-end program to run.

3. **Internet Information Server (IIS)**

   This is the web server used to run the web site.

University of Malaya

Client Machine

1. **Windows 98/NT**

   This is the operating system or platform used in the client machine to run the system.

2. **Internet Explorer 4.0 or higher version**

   The client must use Internet Explorer browser to view the system interface.

University of Malaya

5.2 System Implementation

Implementation is a process comprises of system design structure to a computer readable system. The system will be evolved from scratch design to a run able application. The implementation of this system includes two types of implementation. The first type is the interface implementation, which is the implementation of the front-end of the system. Meanwhile, the second type of implementation is the neural network implementation in the system, which operates as the back-end of the system. Both of these implementations will be explained in the following sections.

5.3 Interfaces and Database Implementation

The interface of this system are developed using HTML and ASP while the database is implemented using Microsoft Access 2000. Each of these two implementations will be discussed below:

5.3.1 Interfaces Implementation

To develop a good interface, a few criteria must be taken into account. Some of the criteria are simple, user-friendly, interactive and so on. The interface that was created has fulfilled these criteria and the example of the interface can be found in Appendix A. The method for implementing this interface is described below:

University of Malaya

Search Form Design

- The search form is used to give user to input their requirement for the property that they want to find. There are five fields in the form, which the users have to input. These fields refer to house type, area, price range, number of bedroom and number of bathroom.
- All these five fields are predefined, where users just have to choose from the drop-down menu to fill in the form. This is to ensure that the input data is standard for all users and thus the search will be simpler.
- User must input all the fields in the form before they can proceed with the search. This feature is to make sure all input fields are input because the search is based on all these five fields. In one of it is not input, then the search cannot be carried on.
- A reset button is provided to help users to reset their form if they have mistakenly input some incorrect choice. It is easier to click a button than go to erase the fields one by one.
- The link to other page of this site can be found on every page of the site. This is to enable the user to access other page faster and easier.

Feedback Form Design

- The feedback form is designed to give users the opportunity to give their opinion on the system. Thus this form will provide a text area for user to input their comments.

University of Malaya

- The fields for users to input their name and e-mail address are also included in this form so that the administrator can contact them.

5.3.2 Database Implementation

This system used the Microsoft Access 2000 database as the DBMS. There are quite a lot of table created to store data for the Filtering Module. Currently, there are four tables created for the Filtering Module. Each of these will be shown below. The following diagram shows the flow between the database and the module.

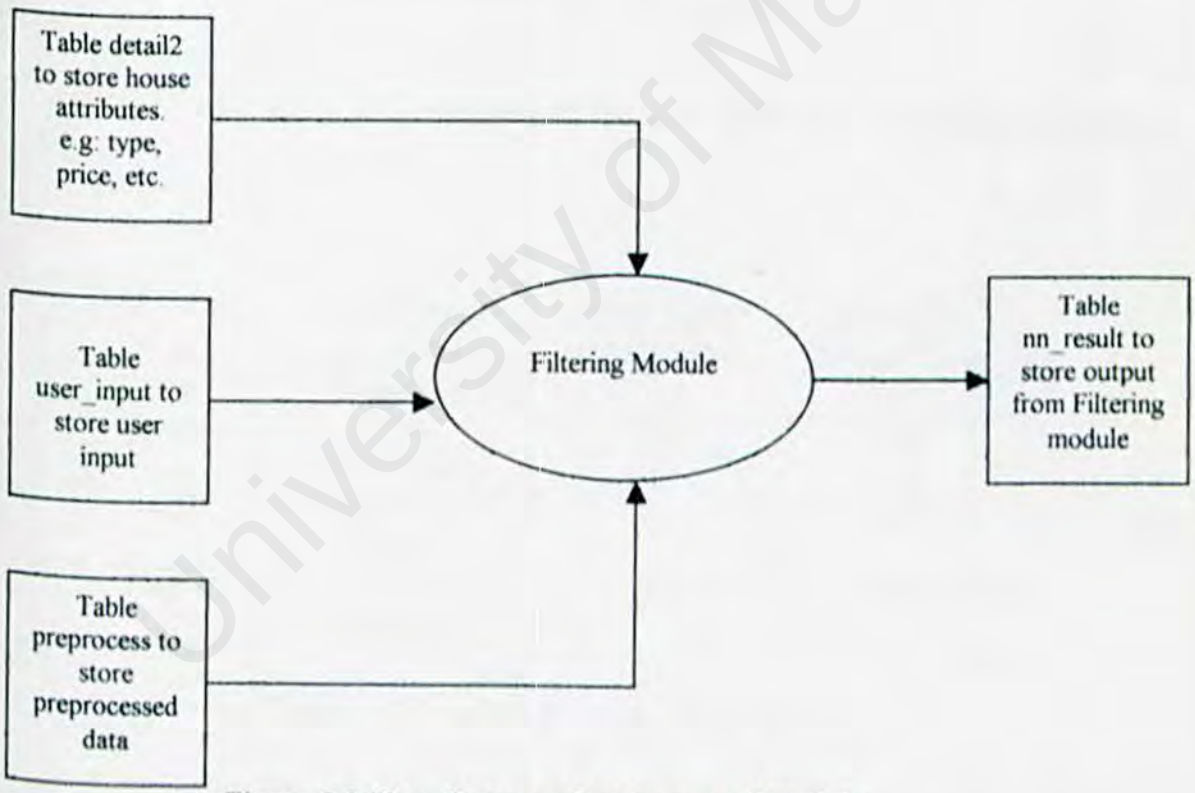


Figure 5.1 Flow Between Database and Module

University of Malaya

- Table *detail2* is used to stored the attributes of a certain house such as house type, area, price, number of bedroom and bathroom, its URL and so on. The table properties and its data type is shown in the table below:

Field Name	Description	Data Type	Field Size
Num	Number	Auto Number	Long Integer
Type	House type	Text	255
Price	House price	Text	255
Area	State	Text	255
URL	House URL link	Text	255
Name	Residential name	Text	255
Location	Specific location	Text	255
Bedroom	Number of bedroom	Text	255
Bathroom	Number of bathroom	Text	255

Table 5.2 Property for Table *detail2*

- Table *user_input* is used to stored the user input and its property is shown below:

Field Name	Description	Data Type	Field Size
Num	Number	AutoNumber	Integer Long
Type	House type	Text	255
Area	Area	Text	255
Price	House price	Text	255
Bed	Number of bedroom	Text	255
Bath	Number of bathroom	Text	255

Table 5.3 Property for Table *user_input*

- Table *preprocessed* is used to store the preprocessed data before it is sent to the neural network for filtering. The properties of this table is shown below:

Field Name	Description	Data Type	Field Size
Num	Number	AutoNumber	Long Integer
Type	House type	Number	Long Integer
Area	Area	Number	Long Integer
Price	House price	Number	Long Integer
Bed	Number of bedroom	Number	Long Integer
Bath	Number of bathroom	Number	Long Integer

Table 5.4 Property for Table *preprocessed*

- Table *nn_result* is used to store the output from the neural network. The properties of this table will be shown below:

Field Name	Description	Data Type	Field Size
Class Type	Class from neural network	Number	Long Integer
ID	Key related to table *detail2*	Number	Long Integer

Table 5.5 Property for Table *nn_result*

University of Malaya

5.4 Neural Network Implementation

The neural network is implemented as the back-end of the system. There are quite a lot of steps required to implement this network. The steps are described below:

Define the setup for neural network

Before a neural network is built, the required parameters, network architecture must be defined first to get a clear view of the neural network that will be implemented later on.

Create training data

After the network is defined, the neural network must be trained. Therefore, a data must be created to train the network. There must be two sets of data namely, train set and test set. The train set is used for training the network and the test set is used to test the network after the training. A set of data must consist of two elements, which are the input vectors and the desired outputs.

Train network

In this step, the network is ready for training and the train set just created will be used here. During the training, the network may not learned what it is supposed to do, thus the learning parameters must be adjusted throughout the training process.

Simulate the neural network

When the network has been trained, it needs to be test to see whether it has learned or not. Here, the test set will be used to test the trained network. If the simulation does not

University of Malaya

give the desired output, the network has to be trained again with different parameters. Then, the network will be tested again.

5.4.1 Preprocessing of Input Vectors

Preprocessing is a process used to convert raw data that is collected into data that will be accepted as an input vectors to the network. The preprocessing of input vectors for this neural network is described as follow:

Collection of raw data

- The raw data meant here is the data in its original form as it is collected from the source.
- Here, the inputs used are the house type, price, area, number of bedroom and number of bathroom. Each of these attributes will be assigned a key, which is a number to identify them. These keys are the raw data. The key for each of the attributes are shown below:

University of Malaya

Num	House Type	Key Assigned
1.	Bungalow	1
2.	Semi-Detached	2
3.	Condominium	3
4.	Double-Storey Terrace	4
5.	Single-Storey Terrace	5
6.	Apartment	6
7.	Town House	7
8.	Flat	8

Table 5.6 Key assigned for house type

University of Malaya

Num	Price Range	Key Assigned
1.	< 99,999	1
2.	100,000 – 199,999	2
3.	200,000-299,999	3
4.	300,000-399,999	4
5.	400,000-499,999	5
6.	500,000-599,999	6
7.	600,000-699,999	7
8.	700,000-799,999	8
9.	800,000-899,999	9
10.	900,000-999,999	10
11.	1,000,000-1,999,999	11
12.	2,000,000-2,999,999	12
13.	>3,000,000	13

Table 5.7 Key assigned for house price

University of Malaya

Num	Area	Key Assigned
1.	Johore	1
2.	Kedah	2
3.	Kelantan	3
4.	Kuala Lumpur	4
5.	Malacca	5
6.	Negeri Sembilan	6
7.	Pahang	7
8.	Perak	8
9.	Perlis	9
10.	Penang	10
11.	Sabah	11
12.	Sarawak	12
13.	Selangor	13
14.	Terengganu	14

Table 5.8 Key assigned for Area

University of Malaya

Num	Number of Bedroom	Key Assigned	Number of Bathroom	Key Assigned
1.	1	1	1	1
2.	2	2	2	2
3.	3	3	3	3
4.	4	4	4	4
5.	5	5	5	5
6.	6	6	6	6
7.	7	7	7	7
8.	8	8	8	8
9.	9	9	9	9
10.	10	10	10	10

Table 5.9 Key assigned for bedroom and bathroom

University of Malaya

Conversion of Data to Input Vectors

- User input data, which consists of house type, price, area, number of bedroom and number of bathroom will be converted into a key according to the above tables.
- The data in the database, which also consists of house type, price, area, number of bedroom and number of bathroom for each residential property will be converted also to their particular key.
- After the conversion, the key from the user input and the key from the database will be compared to see whether it is the same or not. The method used to compare is to get the difference between these two keys.
- This difference between the keys will be stored in the database and it will be retrieved out as the input vectors for the neural network. One set of data consists of type, price, area, bedrooms and bathrooms. Basically, there is only one set of data of user inputs, but there is more than one set of data in the database. Thus, the comparison is made where the users data is compared to every set of data in the database. The next figure shown this:

University of Malaya

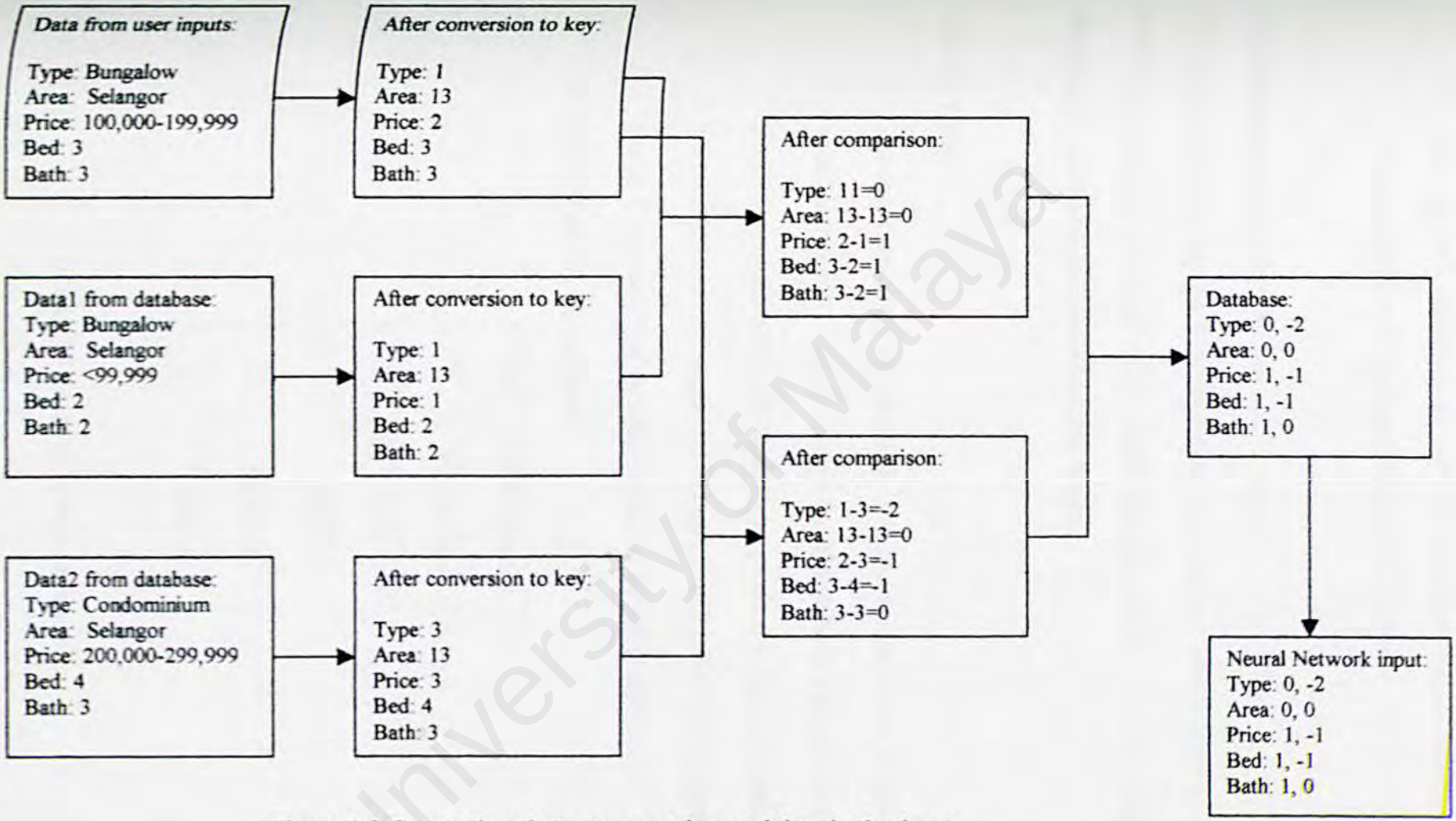


Figure 5.2 Comparison between user data and data in database

University of Malaya

- After the data has gone through the preprocessing, it will be stored in the database and fetch to the neural network as inputs.

5.4.2 Determining Network Target and Output

After the input data has been preprocessed, the network target must be set so that a particular input will give a particular target (desired output). The following section will discussed about the way to determined the target.

Network Target

- The input vector of this network is in matrix form. It is in a 5-by-Q matrix, which means that it has five rows and Q columns. Each of the rows refers to the five input elements, which are type, price, area, bed and bath. The order of the rows here is very important because it will affect the determination of the target. The example of training set can be found in Appendix F.
- The row orders are as follow:

Row	Input Element
1	House Type
2	Area
3	Price
4	Bedroom
5	Bathroom

Table 5.10 Rows order of input vectors

University of Malaya

- The Q columns represent the number of training pattern users want to input. It means that how many set of input vectors will be fed for training. A column represents a particular data of a property in the database. It consists of the house type, area, price, bedroom and bathroom.
- The target output will be classified into two classes, which is represented by the value '1' and '2'.
- The determination of target output, or more precisely, how to put the input vectors into these two classes is based on certain rules. If the value of the input vectors is '0' then it means that this element is the exact match with what the users want. Otherwise, if it is not '0', then it is not a match.
- Here, if the input is classified as class '2', it means that this input is matched with the users requirement. In other words, it is relevant to users. If, on the other hand, the input is classified as class '1', then this input is not relevant to users.
- The criteria used to determine the target output or which classes this input belongs to is to use the number of '0' in a column and the which rows the '0' in. If there are more than three '0' in a column and the first three rows are '0', then this column or input will be classified as class '2', which is relevant to user. Meanwhile, if a column does not fulfill the above criteria, it will be classified as class '1', which means it is not relevant to users. The table below shows some example of the classification.

University of Malaya

	Input 1	Input 2	Input 3	Input 4	Input 5
Type	0	0	3	0	0
Area	0	0	0	0	0
Price	0	0	0	0	1
Bedroom	0	8	0	7	0
Bathroom	0	0	9	5	8
Target	2	2	1	2	1

Table 5.11 Example of input vectors with its targets.

- The rationale for this way to determine the classification is that the first three rows represent the house type, area and price. If these three rows are '0', it means that this column (represent a particular house in the database) match with what the users want to find. These three fields have more priority than the bedroom and bathroom. Thus, even though the bedroom and bathroom rows are not '0', this column is still classified as class '2' as long as the first three rows are '0'.

5.4.3 Neural Network Setup

In order to setup a neural network, all the necessary training parameters and weights must be initialized first before continue with the training. The methods for determining the initial training parameters and weights for the network are described below:

University of Malaya

Training Parameters

- The parameters used for the LVQ network are training function, number of epochs and learning rate.
- The training function used in the network is 'trainwb1' which is the default training function provided by Matlab for LVQ network.
- The initial values of epochs for the network is set to 5000 epochs because this value will be adjusted later on when the training begins.
- The default value for the learning rate in the LVQ network is 0.01.

Transfer Functions

- The LVQ network is a two layers network, thus it has two transfer functions each for one layer.
- Transfer functions for the first layer is *compet* transfer, which is provided by Matlab.
- The second layer uses the *purelin* transfer function.

Weights

- Before beginning with the training, the weights must be initialized to some values first.
- The first layer of the LVQ network will be initialized to the midpoint according to the range of each input. Example: If input is in the range of 0.00-10.00, then the weights will be initialized to 5.00.

University of Malaya

- For the second layer, the layer weights are set so that each output neuron i has unit weights coming to it from class percentages (i) percent of the hidden neurons.

Hidden Neurons

- The hidden neuron is set to five neurons because there are five different elements of input. This value will be adjusted during training.

Input Vectors

- The input vectors are set with five element inputs based on house type, price, area, bedroom and bathroom.
- The range of each element is set according to the maximum value of the element. Example: House type can has eight different values, thus the range will be set to 0-8.
- Below is the range of each element:

Input Element	Range
Type	0 – 8
Price	0 – 13
Area	0 – 14
Bedroom	0 – 10
Bathroom	0 – 10

Table 5.12 Input Range for LVQ Network

University of Malaya

5.4.4 Neural Network Training

When the neural network is set with all the necessary attributes and the training data is ready, then it can be trained to learn what it is supposed to do. Training a network is a difficult task because sometimes the network might not learned at all or it can be over learned. Basically, the training parameters and other attributes of the network determined the learning process. The method of training this neural network is described below:

Training Process

- The training begins when the network attribute is initialized with all the values just described. The training set, which complete with inputs and its desired output is feed into the network in order to get it learned to produce the desired output when the input is provided.
- After the network is trained for the determined epochs, then the input from the test set will be fed into the trained network for simulation. Here, the network weights are set to the weights from the training.
- This time only the input from the test set is fed into the network. The output will not be fed into it. If the output from the simulation is the same with the desired output in the test set, then the network is said to have learned. But if it is different from the desired output from the test set, then the network has not learned and has to be trained again.
- If the network is trained again, then the training parameters have to be adjusted again. The methods for adjusting these parameters are shown in the sections.

University of Malaya

Adjusting Training Parameters

1. Epochs

The method for determining the best number of epochs for training the network is to set the epochs to a certain value. Then, decreased the epochs using a certain value until it reached a small value. The performance for each set of the decreasing epochs is compared and the one that provides the best performance is chosen. After decreased it, the epochs have to be increased from the initial epochs using a certain value to a big value.

Just like before, the epochs, which give the best performance is chosen. Then this epoch will be compared with the chosen epochs from the decreased epochs. The one, which give the best performance, is chosen as the epoch to train the network. This method is used, because determining an epoch to train a network is quite difficult where sometimes a network will be under train if the epoch is too small and the network will be over train if the epoch is too big.

Thus, testing the network with decreasing and increasing number of epochs from a certain value can ensure that all value of the epochs are tested and the one, which give the best performance can be found.

2. Learning Rate

The learning rate is a parameter that is used to control the speed or rate of the network learning. The higher the learning rate, the faster the network learns while for the lower

University of Malaya

the learning rate, the slower the network learns. Although, a higher learning rate can helps the network to learn faster, but sometimes this will make the learning not thorough and the network not robust. When the learning is not thorough, some of the inputs will not generate the correct outputs. On the other hand, if the learning rate is small, the network learning will become very slow but the learning process will be quite thorough.

After considered all the criteria stated above, the method to determine the optimum learning rate for this network is to use the trial and error method. An initial value of the learning rate is set and this value is adjusted by decreasing and increasing it. The value, which gives the best performance to the network, will be used.

3. Hidden Neurons

Initially, the network is set with five neurons. This value will be adjusted during the training period to obtain the optimum hidden neurons. The number of hidden neurons was chosen heuristically.

When the optimum training parameters are chosen, the network will be trained with these parameters. Then the weights will be adjusted according to the LVQ learning rule, which is provided by Matlab in the *'learnlv2'* function.

University of Malaya

5.4.5 Neural Network Simulation

When the neural network has been trained and the correct weights are obtained, it will be used to do the simulation. In other words, the network can now be used to do simulation. Before the simulation, there are a few steps involved to prepare the network for simulation.

Simulates Network

- A network will be initialized with the trained network weights.
- The input vectors will be fed into the trained network and the output will be generated. Here, the network will classify the input vectors into two classes as described earlier.

University of Malaya

5.5 Testing

Software testing refers to verification and validation of the program coded to solve the problem. Verification involves ensuring that the characteristics of a good design are incorporated into the program and the system is actually operates in the way it is expected to be. On the other hand, validation is used to test the execution of the program and system meets the requirements.

The major focus of testing is to find the faults within the program that are not realized at the time of coding. A test is successful only when a fault is discovered or when a failure has been come across. Testing actually involves the iteration of the process of fault identification and fault correction or removal.

In developing large system, testing usually involves several stages. These stages are module testing, integration testing, function testing, performance testing, acceptance testing and installation testing. Each of these stages will be discussed individually.

5.5.1 Module Testing/Unit Testing

The module testing is a way to verify that the small unit function properly with the types of input expected from the design. It has been carried out under a controlled environment where a predetermined set of data has been provided to the modules. In other words, this kind of testing is used to observe what input and its related output actions as well as the data produced. [12]

University of Malaya

In module testing, each of the sub-modules is tested separately. After that, each of the modules will be tested in turn for the creation of the user interface, the input data handling and output to data files, reset and exit from the modules to make sure those modules do exactly what it were designed to perform. Test cases have been developed to show that the input is properly converted to the desired output.

5.5.2 Integration Testing

After the collections of modules have been tested, the next step is to ensure that the interface among the components are defined and handled properly. Thus, integration testing will be performed to achieve this. Integration testing is the process of verifying that the system modules work together as described in the system and program specification. [12]

In this stage, all the individual sub-modules and modules are integrated and tested to ensure that the interfaces between the sub-modules and modules, modules and the main program are handled properly. Here, all the small modules that are tested are isolated first before they are combined into a functional program in the system and tested together.

For this system, the testing approach that has been applied in the integration testing is the bottom-up integration. Each component at the lowest level of the system hierarchy is tested individually first. Here, each of the sub-modules is tested individually first, then the modules in which comprise of sub-modules are tested in turn. Finally, after the

University of Malaya

integration of the modules into a functional program, the main program is tested to ensure that the system performs its works correctly.

5.5.3 Function Testing

After the integration testing is conducted, the function testing should be carried out to assure that the system has the desired functionality. The function test will evaluate the system to determine if the functions described in the requirement specification are actually performed by the integrated system. [12]

5.5.4 Other Testing

The remaining testing should be carried out after the functional testing are performance test, acceptance test and installation test.

Performance test compares the integrated components with the non-functional system requirements such as security, accuracy, speed and reliability. Acceptance test is ran by the users of the system to assure them that the system they wanted was the system that was built for them. The installation test allows users to exercise the system functions and document additional problem that result from being at the actual site.

University of Malaya

5.5.5 Intelligence Testing

The intelligence of this system is embedded in the neural network. Thus, to test whether the system is acting intelligently, the performance of the neural network to perform what it has been taught to learn is tested. The following section discusses the methods taken to test the intelligence of the system:

- Initially, the neural network is tested after it has gone through the training procedure. When the network is trained, a few inputs will be used to test the network.
- Then the neural network will be linked to the system's interface, where the neural network is tested again. This time the network will be tested, as it will be used when the whole system is developed.
- At this step, user will input what they want to find in the search property form. After the form is submitted, the neural network will be call to perform the classification of the data, to classify which one is relevant to user.
- In the search property form, there are five fields, which the user must fill in. The fields are house type, area, price, bedroom and bathroom. If the search results returned by the neural network contains at least three fields that are the same with what the users enters in the search form, then the network has achieved its goal. In other words, the system is considered as intelligent, where it can search for what the users want.

University of Malaya

The following chapter will discuss about the system evaluation and the overall conclusion. Here, the system strength, limitations and future enhancements of the system are presented. At the same time, the problem encountered and knowledge gained during the project will be described also.

University of Malaya

CHAPTER VI

SYSTEM EVALUATION AND CONCLUSION

The final chapter of this system will evaluate the built system in a few areas. It will cover the system strengths and limitations while a few suggestions will be made to enhance the system in the future. In the next sections, the problems encountered and knowledge gained during the entire project. To conclude this chapter, an overall conclusion will be made.

6.1 System Strength

The system strength of this system will be stated below, where the system strong sides are shown.

1. **Simple Interface**

   The interface created for this system is quite simple in design. The purpose of this is to give the user the simplicity of the system. Furthermore, user is also able to fill in the form easily because the design of the form is not complex.

2. **Intelligent Filtering**

   This system will perform an intelligent search in the database, where the relevant information will be filtered out and display to the user. This filtering is using the neural network to filter the relevant data from the irrelevant data. After that, the relevant data, which is required by the users, will be displayed to the users.

University of Malaya

3. Prediction of Residential Property Prices

Unlike other property sites, which just provide the searching for residential property, this system also provides the prediction of the current property prices on the market. This feature will assist user in making decision about which property to buy from the result of the search.

University of Malaya

6.2 System Limitations

System limitations are the shortcomings of the system that was built. Each of the limitations found will be discussed as follow:

1. **Offline Searching**

   This system will search for the result offline, where it get the data from database and then send it to the neural network to perform the searching for relevant data. Thus, the database has to be updated often so that new data is available for searching.

2. **Search Space**

   This system is built to search for residential houses that are available in Malaysia only. Therefore, the system can only be used to search for houses in Malaysia according to state.

3. **Browser Limitations**

   Currently, this system can only supports Internet Explorer 4.0 or higher version. Thus, the client must use Internet Explorer only to browse the interface.

University of Malaya

6.3 Future Enhancement

It is found out that the system developed has some deficiencies. Thus, it is suggested here, that some enhancement be made to the system in the future to make it more powerful. Below are some of the suggestions for the enhancement:

1. **Integrate with a more powerful database**

   The database used in the system is Microsoft Access 2000. It is possible that the system will be integrated with a more powerful database such as Microsoft SQL server 6.5 or Oracle and so on. A powerful database can provide flexible handling of data and systematic arrangement in the database. In addition to that, this kind of database is easier to manipulate existing data such as add, delete, modify and it also can supports a lot more users.

2. **Online Searching**

   Currently, the system only performed offline searching from the database. In the future, the system can be enhanced to perform online search from the WWW. With online searching, the data that is accumulated will be stored in the database and then feed into the neural network to perform the filtering of relevant data for users. Thus, the data provides to the users will always be the latest because it is found from the WWW.

University of Malaya

and thus they can train their own network, as they like without having to consider about other users.

6. **User manual for learning and using neural network**

If users are provided a neural network as said, then they must have some basic knowledge about neural network in order to train it. Thus, a user manual to learn and use a neural network in the context of this system will be provided to users. With this user manual, users can have a better view of how the neural network works in this system and so they can train it the right way.

7. **Develops a stand-alone applications using Matlab**

This system is currently using HTML and ASP in developing its front-end and also uses IIS as its web server. Thus, it needs to be linked to Matlab using a DLL file because the environment is different. This will makes the system slow down because it has to invoke Matlab every time the system is used.

Fortunately, Matlab has its own web server and users can also create their own interface using the graphics properties. With this ability, this system can be enhanced to a stand-alone application using only Matlab to develop it. This will make the system flow better and increased its speed because the front-end and back-end of the system are using the same development environment.

University of Malaya

3. **Integrate with other browsers**

The system use only the Internet Explorer browser, thus, in the future it is hope that it can use other browsers such as Netscape or other browsers. If this is done, users can use this system more flexible where they can use any browser to browse it.

4. **Provide user the ability to train the filtering agent**

The filtering agent, which is the neural network for filter out the relevant data, is trained and updated by the administrator for now. This feature can be extended, where users can train the network time by time so that the network performance will get better. Furthermore, it will save the administrator a lot of time to retrain the network frequently because if the network is not retrain, the network performance will go down as time goes by. If the users train the network frequently, the network will give a more intelligence performance.

5. **User login account**

For now, this system does not provide user login and thus concurrency problem will arise sometimes. In order to solve this problem, the system will provide a user login feature, where users can have their own account. In the account, users are provided the history of their previous activity. Previously, it is mentioned that users can train the network themselves, so together with this feature; users can have their own neural network for filtering. In other words, whenever a user registered for an account, a pre trained neural network will be provided to them

University of Malaya

6.4 Problems Encountered And Solutions Taken

During the entire course of this project, there are a lot of problems encountered whether non-technical or technical problems. Fortunately, solutions are found for every problem that was encountered. The following are the problems encountered and the solutions taken to overcome it.

1. **Problem: Lack of knowledge in Artificial Neural Network**

   Artificial neural network is a new subject to me and I have little knowledge about it. Thus, I have to study the subject for the project. This subject proves to be quite difficult to understand especially for those who do not has any prior knowledge of it. This is because there are many new terms in the subject and the way the network function is also hard to understand. Because of this, learning the artificial neural network at first is quite a challenge.

   Solution Taken:

   The solution I have taken to solve this problem is that I do a lot of research on the subject. I also sought some advice from my project supervisor whenever I encountered some difficulties in understanding the subject.

2. **Problem: Difficulty in choosing a neural network to use**

   There are many types of neural network available and each of them has its own use. Furthermore, the learning rules and the training method are also different for

University of Malaya

each network. Thus, it is very difficult to choose what types of network to implement in this system so that it is suitable for the system.

Solution Taken:

I have done a lot of researching in order to determine which type of network to implement in the system. Basically, the determination of a network is based on the problem to be solved. Different type of network can solve different problem like classification, pattern recognition and so on. Thus, I have to determine the problem to be solved by the network first before I can choose the type of network to be used. Finally, the LVQ network is chosen because it is suitable to do classification define by the users, which is needed by the system.

3. **Problem: Inexperience in the chosen programming languages**

During the coding period, I have encountered a lot of problems with the programming languages. This is due to my lack of experience in the programming languages such as VBScript, Matlab and so on. Therefore, when coding I have made a lot of errors and thus made problems sufficed in the system.

Solution Taken:

To solve this problem, I have done some trials and errors on the coding to make it work. Besides this, I also referred to reference books whenever it is possible. I also get some information from some of my friend regarding the problem because some of them already have some knowledge of the programming languages.

University of Malaya

4. **Problem: Difficulties in training the neural network**

To train a neural network to do what you want desired is not an easy task. This is what I experienced during the training of the neural network. Whenever I tried to train it, the result it gave is not what I want. Therefore, I have to train it many times with different parameters because the parameters can affect the training. The parameters that meant here are number of epochs, learning rate, learning function, number of neurons and so on. The LVQ network is quite difficult to train because it has to classify the data as defined by the users and there are quite a lot of data to be classified.

Solution Taken:

In order to achieve the right training, I have done some trials and errors method where I change the parameters until the network get the right training. In other words, the networks get to learn what it is supposed to do.

University of Malaya

6.5 Knowledge Gained

After finishing this project, I have gained a lot of knowledge as well as experience in various aspects. The knowledge gained is as follow:

1. **Knowledge of Artificial Neural Network**

   I have really learned what is an artificial neural network is and the power it can provides to do something. Basically, neural network can be applied in many areas where it can replace conventional programming whenever possible.

   The advantages of neural network compare to conventional programming are many such as neural network can do classification, pattern recognition, which conventional programming still cannot do. With the knowledge of neural network, one can choose which programming method is more suitable for a system. In other words, there are other ways to develop a particular system besides conventional programming.

   I also learned how to apply a neural network to an application. Furthermore, the techniques to train and create a neural network are also learned. Besides all this, I have gained the knowledge of how a LVQ network works and how to use it for classification.

University of Malaya

2. Techniques in designing and maintaining a database

There are many ways to design and maintain a database. But to choose the right way needs some experience in this area. This is what I learned during the implementation of the database of this system. It can be said that to design a database is quite a difficult task because a system needs a lot of table and queries and the relationships between tables also have to be declared. This task needs a lot of effort to do especially when I do not have experience in it. But after a while dealing with it, I start to get the control of it and the task have became easier.

The database needs to be maintained after it is created such as updating new information or delete irrelevant information. The techniques for this task is also not easy but with some experience it can be done. Therefore, after this project, I have learned how to design and maintain a database to make the system more manageable.

3. Knowledge of using Matlab

Matlab is a powerful programming language for technical computing. It is used in this system to implement the neural network and provide access to database. After using it, I really have learned the benefits it can give to create an application. It can create a stand-alone applications and also can performed a lot of technical computing in the engineering field like signal processing and

University of Malaya

so on. Matlab provide its own web server for web applications and it also can convert an M-file to C/C++ source code using the provided C/C++ compiler.

University of Malaya

6.6 Conclusions

After conducting some analysis and survey on the project previously, it can be concluded that an intelligent agent is built to solve problem where conventional system still lacks the ability to do so. Intelligent agent can be applied in many areas, but it is most useful in electronic commerce especially in online transaction. This agent will help user to search for relevant information on certain products. Housing development is an area where this system can be applied to assist user to make decision on what house they want to buy.

Finally, the system that was proposed in the 3181 reports has been built with a little modification to the previous proposed system. The modifications made are the searching part, where the searching is performed on the offline rather than online as proposed previously. This modification does not affect the performance of the system, because the system is built to do intelligent task like providing users with relevant information, which is done in the neural network. However, it is hoped that this features will be implemented in the future.

The system built, can do two main tasks, which are the searching of property and prediction of the current property prices. These two functions are implemented using the neural network method. The searching of property, which is the Filtering Module discussed earlier can find the most relevant property to users according to their requirement. LVQ network is used to implement this function.

University of Malaya

During the progress of this entire project, I have learned a lot of new things and also have encountered a lot of problems, which were mentioned in the above sections. The most difficult part in the project is the coding of the system. This is because, new programming languages have to be learned and a lot of errors found during the coding. Besides the technical things, I also have the opportunity to discover about the importance of teamwork since this is a group project.

This system can now be used to assist users in making decision when they want to buy residential properties. On the whole, the objectives of this system is met, where users is provided relevant information on what they want as well as a price predictor to assist them in making decision when they are buying houses.

University of Malaya

BIBLIOGRAPHY

[1] Elaine Rich and Kevin Knight, *Artificial Intelligence*, McGraw-Hill, Inc., 1991, pp. 3.

[2] Mohamed H. Hassoun, *Fundamental of Artificial Neural Network*, Bradford Book, 1995, pp.1-9.

[3] J.G. Taylor, *The Promise of Neural Network*, Springer-Verlag, 1993, pp. 9-11.

[4] Phillip D. Wasserman, *Advanced Methods in Neural Computing*, Van Nostrand Reinhold, 1993, pp.1-5.

[5] PC-AI Genetic Algorithm, http://www.pcai.com/pcai/New_Home_Page/ai_info/genetic_algorithms.html, Date Referred: 23 June 2000.

[6] Using an Intelligent Agent to Enhance Search Engine Performance, http://www.firstmonday.dk/issues/issue2_3/jansen/index.html, Date Referred: 17 July 2000.

University of Malaya

[7] Electronic Commerce, http://www.ecentre.org.uk/frameset_businesstech.asp, Date Referred: 19 July 2000.

[8] George McDaniel, *IBM Dictionary of Computing*, McGraw-Hill, Inc., 1994, pp. 354-355.

[9] Bargain Finder, http://bf.cstar.ac.com, Date Referred: 10 June 2000.

[10] Firefly, http://www.agents-inc.com, Date Referred: 17 June 2000.

[11] L.P. Shari, *Software Engineering: Theory and Practice*, Prentice Hall International, 1998.

[12] Pfleeger, S.L., *Software Engineering: A Practitioner Approach*, Prentice Hall International, 1998, pp. 48-51, 57, 135-144, 192-195, 244-245, 401-408.

[13] Mastering Series Training Sampler, 1998, http://www.microsoft.com/mastering/sampler/mdad/mv600209.htm, Date Referred: 20 November 2000.

[14] Mara, M.J., *VBScript Sourcebook*, Wiley Company Publishing, Indianapolis, 1996.

[15] Castro, E., *HTML for the World Wide Web*, New York, Peachpit Press, 1996.

University of Malaya

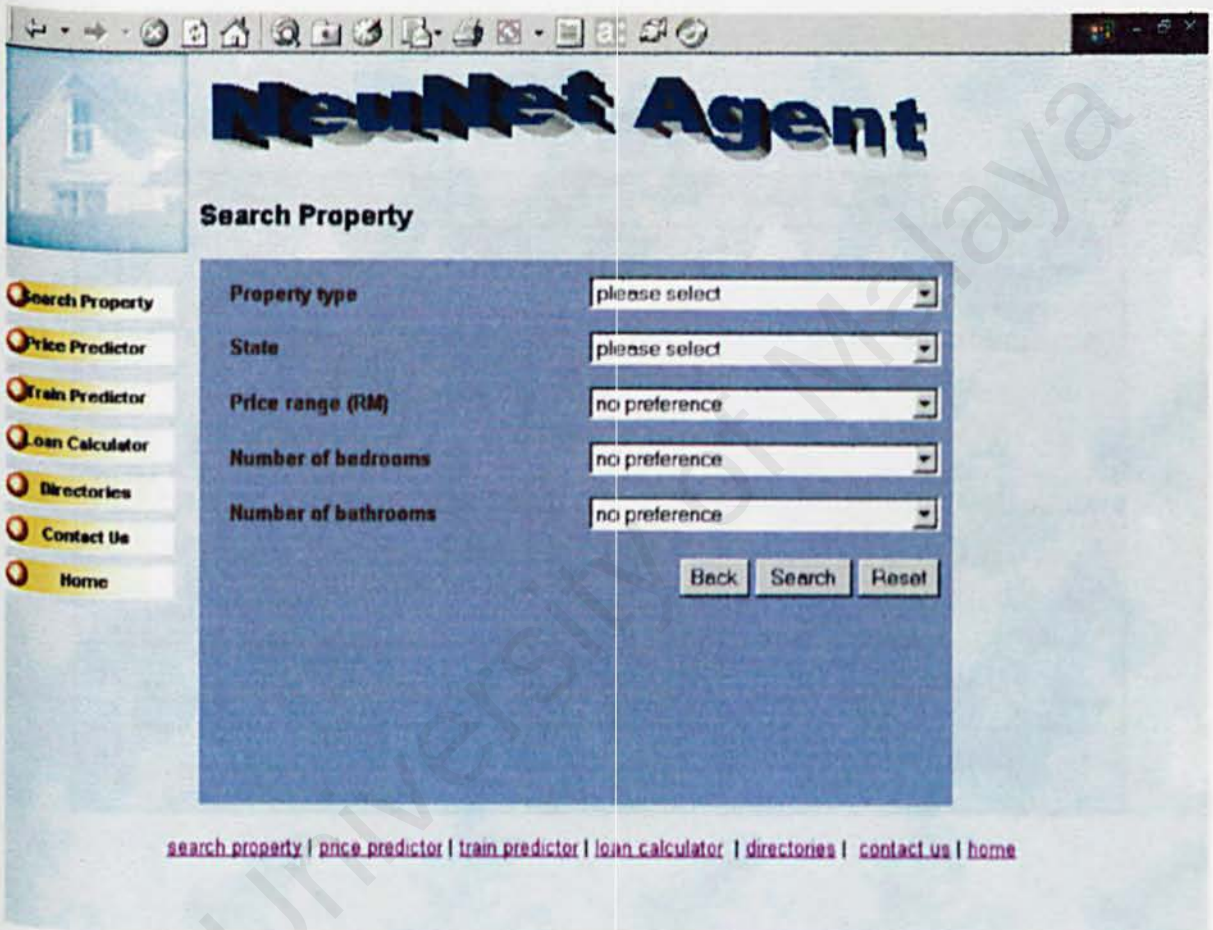
[16] Homer, A., *Professional ASP*, Bermingham, Wrox Press, 1997.

[17] Howard Demulti and Mark Beale, *Neural Network Toolbox User Guide*, Version 3.0, The Mathworks, Inc., 1998, pp. 257-258.

University of Malaya

APPENDICES

Appendix A: User Input Interface



University of Malaya

Appendix B: Result Page

NeuNet Agent

SEARCH RESULT

Num	Residential Name	Type	Area	Location	Price	Bedroom	Bathroom	URL Link
1	N/A	Condominium	Selangor	N/A	RM150000	3	2	http://www.metrohomes.com.my
2	Desa View Condo	Condominium	Selangor	Taman Melawati	RM155000	4	2	http://www.skybusiness.com/apt-s-pg01
3	Kiara Park	Condominium	Selangor	Taman Kiara	RM398000	4	2	http://www.propertylink.com.my
4	Subang Perdana Court 1	Condominium	Selangor	Subang Jaya	RM123000	3	2	http://www.skybusiness.com/apt-s-pg01
5	Istara Apartment	Condominium	Selangor	Petaling Jaya	RM600000	3	2	http://www.eproperty2000.com.my
6	Bandar Sungai Long	Condominium	Selangor	Kajang	RM187000	3	2	http://www.eproperty2000.com.my
7	Sri Mahligai	Condominium	Selangor	Shah Alam	RM250000	4	3	http://www.malaysiadirectory.com/epi-bin/view.cgi

University of Malaya

Appendix C: Attributes of Neural Network in Matlab

```
net =

    Neural Network object:

    architecture:

     numInputs: 1
     numLayers: 2
     biasConnect: [0; 0]
     inputConnect: [1; 0]
     layerConnect: [0 0; 1 0]
     outputConnect: [0 1]
     targetConnect: [0 1]

     numOutputs: 1  (read-only)
     numTargets: 1  (read-only)
     numInputDelays: 0  (read-only)
     numLayerDelays: 0  (read-only)

    subobject structures:

          inputs: {1x1 cell} of inputs
          layers: {2x1 cell} of layers
          outputs: {1x2 cell} containing 1 output
          targets: {1x2 cell} containing 1 target
          biases: {2x1 cell} containing no biases
          inputWeights: {2x1 cell} containing 1 input weight
          layerWeights: {2x2 cell} containing 1 layer weight

    functions:

        adaptFcn: 'adaptwb'
        initFcn: 'initlay'
        performFcn: 'mse'
        trainFcn: 'trainwb1'

    parameters:

        adaptParam: .passes
        initParam: (none)
        performParam: (none)
        trainParam: .epochs, .show, .time, .lr
        weight and bias values:
```

University of Malaya

IW: {2x1 cell} containing 1 input weight matrix
LW: {2x2 cell} containing 1 layer weight matrix
b: {2x1 cell} containing no bias vectors

other:

userdata: (user stuff)

University of Malaya

Appendix D: Example of Network Weights

Input Weights (net.IW{1,1}):

2.6734	1.5755	0.9135	1.7293	1.5996
4.8293	7.6972	9.0835	7.0383	7.3286
-0.1424	-0.2274	-0.1571	3.0965	2.2601
5.7994	7.1682	5.4595	5.8997	6.1521
5.7994	7.1682	5.4595	5.8997	6.1521

Layer Weights (net.LW{2,1}):

1	1	0	0	0
0	0	1	1	1

University of Malaya

Appendix E: Example of Data For Training

Training Set

```
P=[0 0 3 0 7  6 0 0 5 2 0 0 1 0 1 1 0 0 0 0 7 8 0 0 0 0  0 0 8 6  5 1;
   0 2 2 2 0 13 0 9 0 6 0 0 0 0 1 0 2 1 0 0 0 4 0 7 0 0 0  0 0 3 0  6 0;
   0 0 1 0 12 0 0 7 0 0 0 0 0 0 0 0 3 2 0 0 0 8 0 9 0 0 0  0 0 2 0  7 0;
   0 2 9 5 8  0 2 0 0 3 7 0 0 0 0 0 4 3 1 3 0 9 0 3 9 2 3  9 0 7 0  8 1;
   0 0 0 4 9  0 1 0 0 0 8 3 0 0 1 5 4 1 0 0 8 0 3 2 1 10 0 0 1 10 9 1];

T=[2 1 1 2 1  1 1 1 1 1 2 2 1 1 1 1 1 2 2 2 1 1 1 2 2 2  2 2 1 1 1 1 ];
```

P = Input Vectors.

T = Target Output.

Test Set

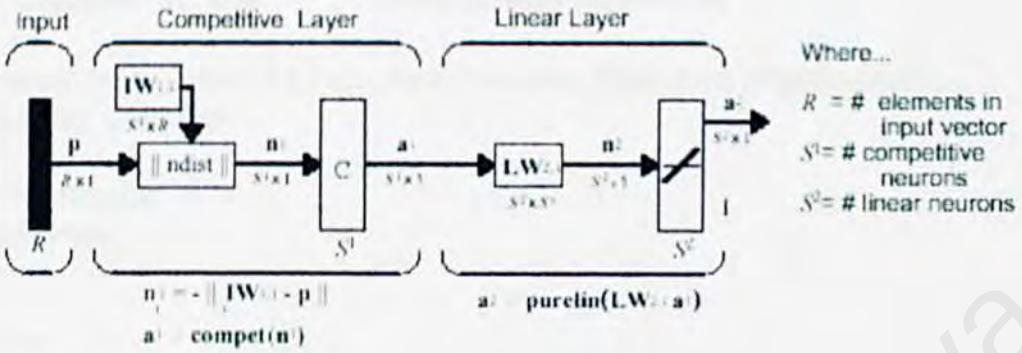
```
P2=[0 3 0 0 0 0 0 0 5 0 0 5 3 3 0 0 5 0 0 2  0 2  0 2 3 3 0 0 3 0  3 ;
    0 0 0 0 0 0 0 2 0 0 2 2 2 0 0 2 2 0 0 0  2 2  2 0 0 2 0 0 0 0  0 ;
    0 1 0 0 0 0 0 0 1 1 0 0 0 1 0 0 1 2 0 11 0 10 8 9 1 0 0 0 0 0  1 ;
    2 1 0 0 1 9 0 3 1 1 1 1 1 1 0 0 1 1 1 0 2  4 3  1 2 1 0 0 0 1 10 0 ;
    1 1 0 0 0 6 0 0 1 1 1 1 1 1 0 0 0 1 0 6 2  0 2  1 1 1 1 3 8 1 0  0];

T2=[2 1 2 2 2 2 2 1 1 1 1 1 1 1 2 1 1 1 2 1  1 1  1 1 1 1 2 2 1 2  1];
```

P2 = Input Vectors

T2 = Target Output

University of Malaya

Appendix F: LVQ network architecture



IW {1,1} = Input Weights.

LW {2,1} = Layer Weights.

C = Competitive Transfer Function (*compet*).

a (1) = Output from competitive layer.

a (2) = Output from linear layer.

n (1) = Input to *compet* transfer function.

n (2) = Input to *purelin* transfer function.

University of Malaya

Appendix G: Sample Matlab Source Code

```
%File simul.m

feature('dispatchjava' , 1); %open database toolbox

conn = database('url_store'," ," );%open database connection

curs = exec(conn , 'select ID,Type,Area,Price,Bed,Bath from preprocessed');
%execute sql statement

curs = fetch(curs);
a2 = curs.Data;


for i=1:97
  id(i) = [a2(i,1)];  %get data
  type(i) = [a2{i,2}];
  area(i) = [a2{i,3}];
  price(i) = [a2{i,4}];
  bed(i) = [a2{i,5}];
  bath(i) = [a2{i,6}];

end


p2 = [type;area;price;bed;bath]; %input of data to be simulate



net = newlvq([0 8;0 14;0 13;0 10;0 10] , 5 , [.6 .4]);% create a LVQ network


net.IW{1,1} = [ 4.5700   7.4209   8.8886   6.8948   7.1908; %initialized input weight
                3.1004   1.5374   0.8177   2.1122   1.5276;
               -0.1922  -0.3279  -0.2823   3.4500   2.5928;
                6.0857   7.4281   5.2091   5.9396   6.4371;
                6.0212   7.4257   5.2477   5.9071   6.4007];


net.LW{2,1} = [1 1 1 0 0 ; %initialized layer weight
               0 0 0 1 1];
```

University of Malaya

```
a = sim(net,p2); %simulate the network
y = vec2ind(a); %change the output from vector to indices

for j=1:48

  exdata{j,1} = id{j}; %get data to be exported
  exdata{j,2} = y(j);

end


cur2 = exec(conn, 'select ID,result from nn_result');
cur2 = fetch(cur2);
r = cur2.Data;

del = exec(conn, 'delete * from nn_result');

colnames = {'ID','result'};

get(conn, 'AutoCommit');


insert(conn, 'nn_result', colnames, exdata);%insert data to database

del2 = exec(conn, 'delete * from nn_result where result=1');

close(conn); %close database connection
```

University of Malaya

Appendix H: Example of ASP Source Code

```
'File datain.asp

<%@ Language=VBScript %>
<HTML>
<HEAD>
<META NAME="GENERATOR" Content="Microsoft Visual Studio 6.0">
</HEAD>
<BODY background=Clear_Day_.jpg>
<FORM action=dis.asp id=form1 method=post name=form1
style="VERTICAL-ALIGN: middle"><STRONG>


<%
 Dim ptype, price , area , bed , bath, ans

 Set ptype = Request.Form("Type")
 Set price = Request.Form("price")
 Set area = Request.Form("state")
 Set bed = Request.Form("bedrm")
 Set bath = Request.Form("bathrm")

 if ptype = "please select" Then
 Response.Write("Please input the house type")
 ElseIf price = "no preference" Then
    Response.Write("Please input the price")
 ElseIf area = "please select" Then
    Response.Write("Please input the state")
```


University of Malaya

```
    ElseIf bed = "no preference" Then
    Response.Write("Please input number of bedroom")
  ElseIf bath = "no preference" Then
    Response.Write("Please input number of bathroom")
  Else
   End if

Dim condb , rsquery , strSQL
Set condb = Server.CreateObject("ADODB.Connection")
condb.ConnectionString = "Provider=Microsoft.Jet.OLEDB.4.0; Data
Source=C:\Intelligent Agent\url_store.mdb
 condb.Open

 Set rsquery = Server.CreateObject("ADODB.Recordset")
 strSQL2 = "DELETE * FROM user_input"
rsquery.Open strSQL2, condb
strSQL = "SELECT * FROM user_input"
rsquery.Open strSQL, condb
str1= "INSERT INTO user_input (type, area,price, bathroom, bedroom)" & _
      "VALUES ('" & ptype & "'," & area & "," & _
      "'" & price & "'," & bath & "'," & bed & "')"
condb.Execute str1
rsquery.Close
condb.Close
%>
```

University of Malaya

```
You have selected these fields:
</STRONG><STRONG>
<br></STRONG><STRONG>Type       &
nbsp;
: <%=ptype%></STRONG><STRONG><br>
 Price Range: <%=price%><br>
 State         
: <%=area%><br>
 Bedroom    : <%=bed%><br>
 Bathroom   : <%=bath%>  <br>


<%
 'Assign a key to user input.
 Dim conn2 , strk1
 Dim user_type, user_area, user_price, user_bathroom, user_bedroom
 Set conn2 = Server.CreateObject("ADODB.Connection")
 conn2.ConnectionString = "Provider=Microsoft.Jet.OLEDB.4.0; Data
  Source=C:\Intelligent Agent\url_store.mdb"
 conn2.Open
 Set del = conn2.Execute("Delete * From user_id")

 Set strk1 = conn2.Execute("Select type,typeID From user_input,type2 Where type =
 House_Type ")

 Set strk2 = conn2.Execute("Select area,areaID From user_input,area2 Where area =
 state ")
```

University of Malaya

```
Set strk3 = conn2.Execute("Select price,priceID From user_input,price2 Where price =
Price_Range ")

Set strk4 = conn2.Execute("Select bathroom,bathID From user_input,bath Where
bathroom = Bath")

Set strk5 = conn2.Execute("Select bedroom,bedID From user_input,bed Where
bedroom = Bed ")
Set ptype = strk1("typeID")
Set parea = strk2("areaID")
Set pprice = strk3("priceID")
Set pbath = strk4("bathID")
Set pbed = strk5("bedID")

Set strk7 = conn2.Execute("INSERT INTO user_id (typeidu, areaidu, priceidu, bedidu,
            bathidu)" & _
           "VALUES ('"& ptype & "'," & parea & "'," & _
           "'" & pprice & "'," & pbed & "'," & pbath & "')")

conn2.Close
%>


<%
Dim conn3 , rsquery2 , strpSQL , strp1, strp2, strp3 , strpSQL2
Set conn3 = Server.CreateObject("ADODB.Connection")
```

University of Malaya

```
conn3.Connectionstring = "Provider=Microsoft.Jet.OLEDB.4.0; Data
Source=C:\Intelligent Agent\url_store.mdb"
conn3.Open

Dim typeid , areaid , priceid , bedid , bathid
 strp2 = conn3.Execute("DELETE * FROM preprocessed")
Set strpSQL = conn3.Execute("SELECT * FROM user_id")

Set typeid = strpSQL("typeidu")
Set areaid = strpSQL("areaidu")
Set priceid = strpSQL("priceidu")
Set bedid = strpSQL("bedidu")
Set bathid = strpSQL("bathidu")

Set strp = conn3.Execute("SELECT * FROM detail2")
do until strp.EOF
Set ID = strp("detailID")
Set htype = strp("House_Type")
Set area = strp("Area")
Set price = strp("price_id")
Set bed = strp("Bedroom")
Set bath = strp("Bathroom")

pro_type = strpSQL("typeidu") - strp("House_Type")
pro_area = strpSQL("areaidu") - strp("Area")
pro_price = strpSQL("priceidu") - strp("price_id")
pro_bed = strpSQL("bedidu") - strp("Bedroom")
```

University of Malaya

```
pro_bath = strpSQL("bathidu") - strp("Bathroom")

if pro_type<0 Then
    pro_type = pro_type*(-1)
 End if
 if pro_area<0 Then
    pro_area = pro_area*(-1)
 End if
 if pro_price<0 Then
    pro_price = pro_price*(-1)
 End if
if pro_bed<0 Then
    pro_bed = pro_bed*(-1)
 End if
if pro_bath<0 Then
    pro_bath = pro_bath*(-1)
 End if
 strp3 = conn3.Execute("INSERT INTO preprocessed (Type, Area, Price, Bath, Bed,
        ID)" & _
        "VALUES ('" & pro_type & "','" & pro_area & "'," & _
        "'" & pro_price & "','" & pro_bath & "','" & pro_bed & "','" & ID & "')")
strp.MoveNext
Loop
conn3.Close

%>
```

University of Malaya

```
</STRONG></FORM>

<FORM action=dis.asp id=form1 method=post name=form1

style="VERTICAL-ALIGN: middle">

<P><STRONG>Please click the submit button to confirm the
selection.</STRONG></P>

<P> </P>

<P><INPUT name=submit1 type=submit value=Submit
>           <input
name="Back" type="button" value="Back"
onClick="javascript:self.history.back()"></P>

<P>

</P></FORM>

<P><LABEL></LABEL> </P>

<P><LABEL></LABEL> </P>

</BODY>

</HTML>
```

University of Malaya

Appendix I: User Manual

Main Page

Links

Search Property

- Function: Go to the Search Property page.

Price Predictor

- Function: Go to the Price Predictor page.

University of Malaya

Train Predictor

- Function: Go to the Train Predictor page, which is accessible only to administrator.

Loan Calculator

- Function: Go to the Loan Calculator page.

Directories

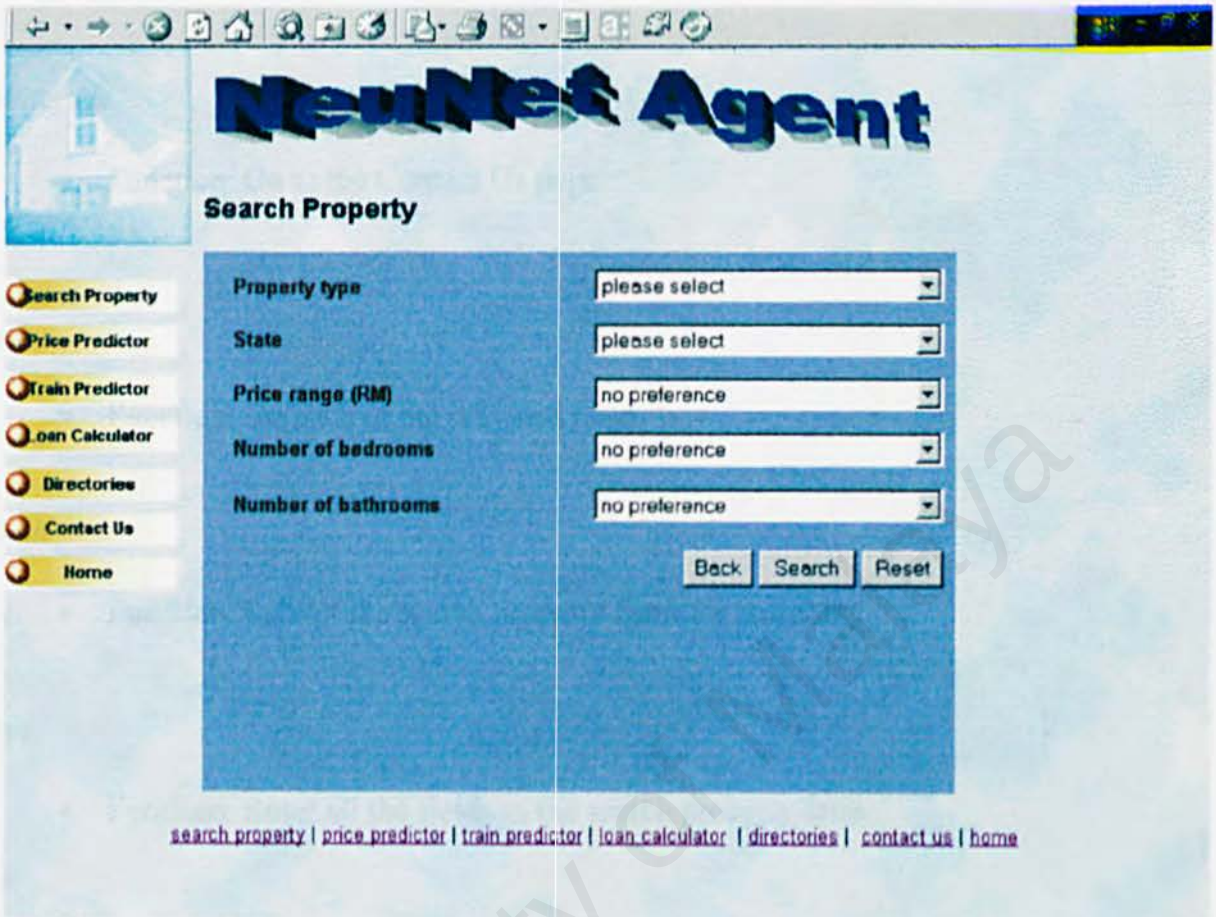
- Function: Go to the Directories page.

Contact Us

- Function: Go to the Contact Us page.

University of Malaya

Search Property Page



Buttons

Search Property

- Function: Go to Search Property page.

Price Predictor

- Function: Go to the price predictor page.

Loan Calculator

- Function: Go to the Loan Calculator page.

Directories

- Function: Go to the Directories page.

Contact Us

- Function: Go to the Contact Us page.

Back

- Function: Go back to the previous page.

Submit

- Function: Submit the search property form for searching.

Reset

- Function: Reset all the fields in the search property form.

Search Property Form

Property Type Field

- Function: Let user select the property type of their search property.

State Field

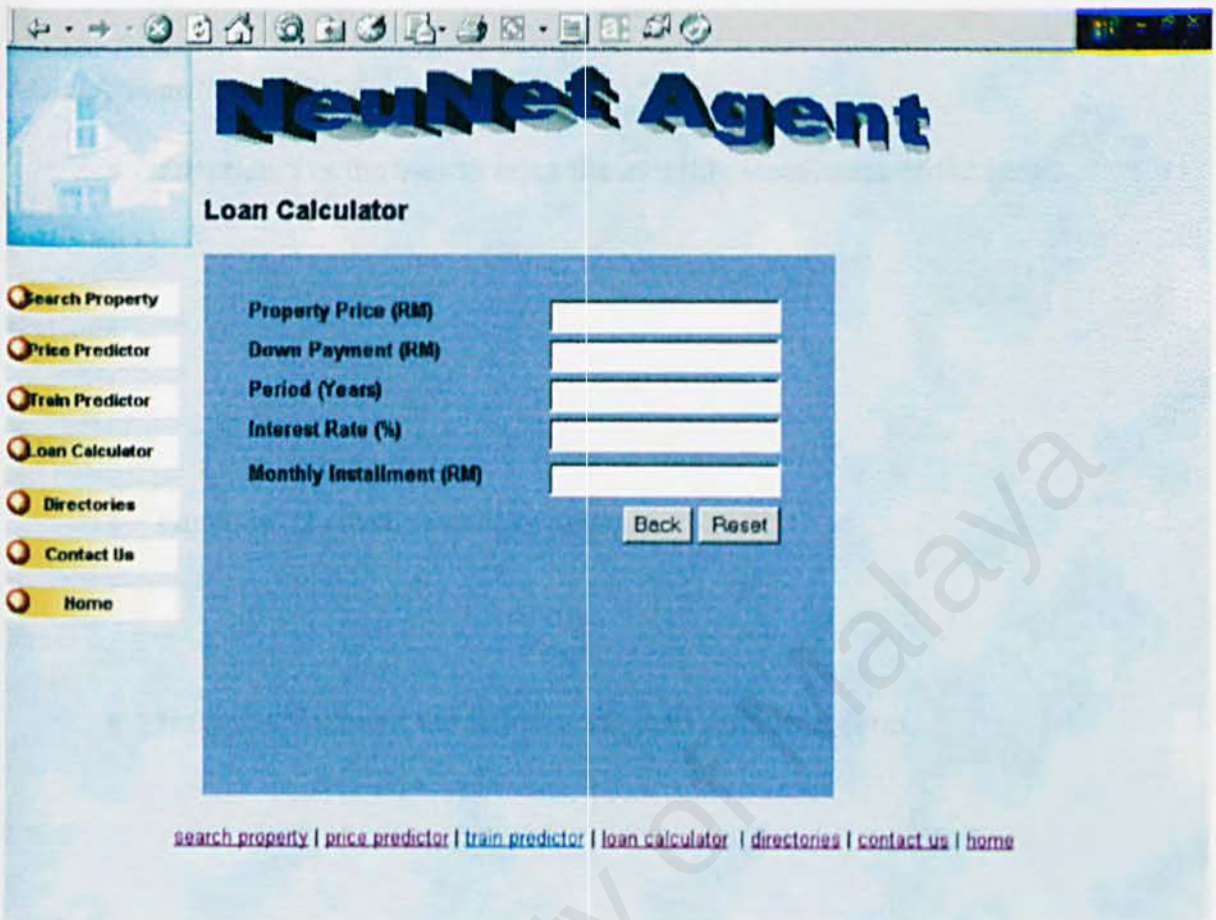
- Function: Let user select the state of their search property.

Price Range

- Function: Let user select the price range of their search property.

University of Malaya

Loan Calculator Page



Loan Calculator Form

Property Price Field

- Function: Let user to input the price of the property they want to buy.

Down Payment Field

- Function: Let user to input the down payment of the property they want to buy.

Periods Field

- Function: Let the user to input the periods of the loan.

University of Malaya

Interest Rate Field

- Function: Let the user to input the interest rate of the loan.

Monthly Installment Field

- Function: Let the user to input the monthly installment of the loan.

Buttons

Back

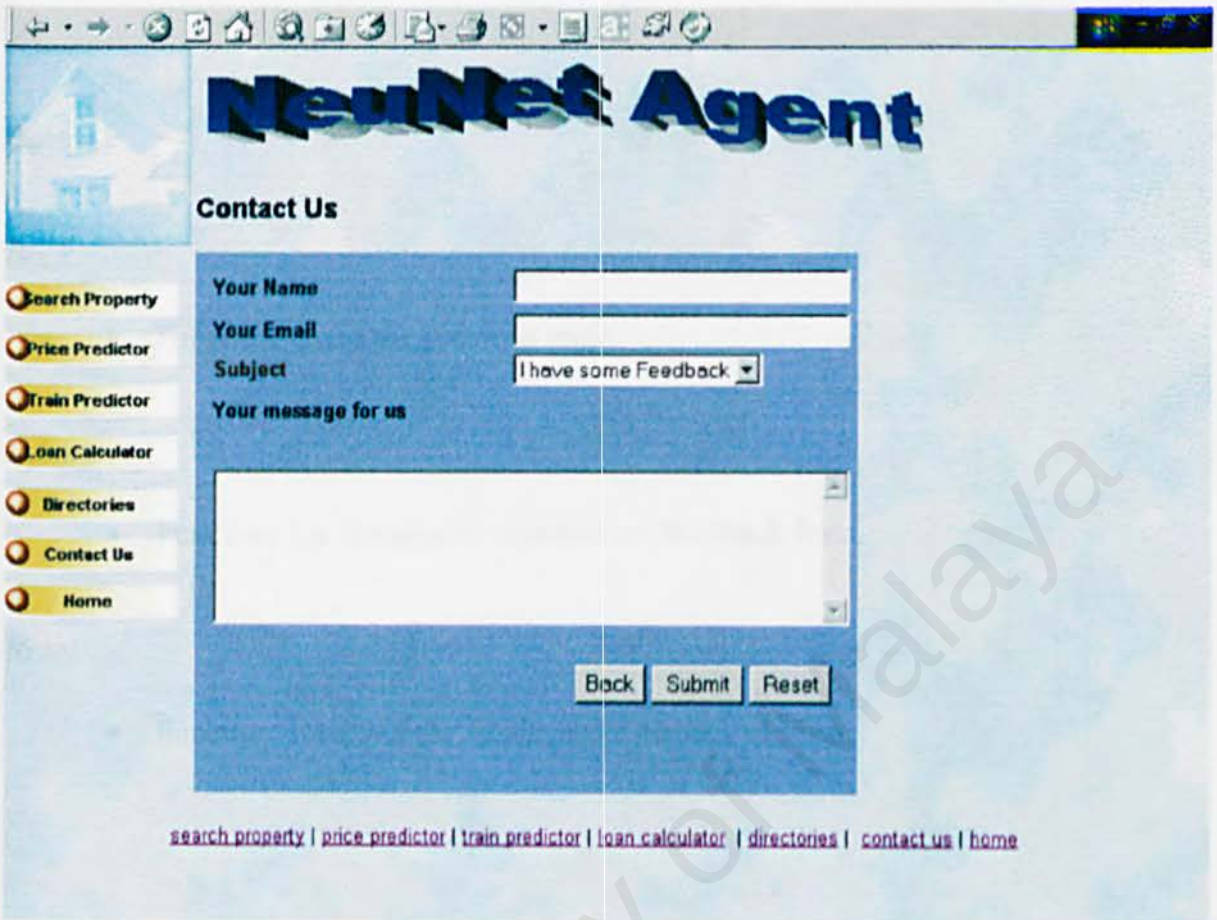
- Function: Go to the previous page.

Reset

- Function: Reset all the fields in the loan calculator form.

University of Malaya

Contact Us Page



Contact Us Form

Your Name Field

- Function: Let the user to input their name.

Your Email Field

- Function: Let the user to input their E-mail address.

Subject Field

- Function: Let the user to select the type of feedback they want to send.

University of Malaya

Your Message For Us Field

- Function: Let the user to input the feedback to the administrator.

Buttons

Back

- Function: Go to the previous page.

Submit

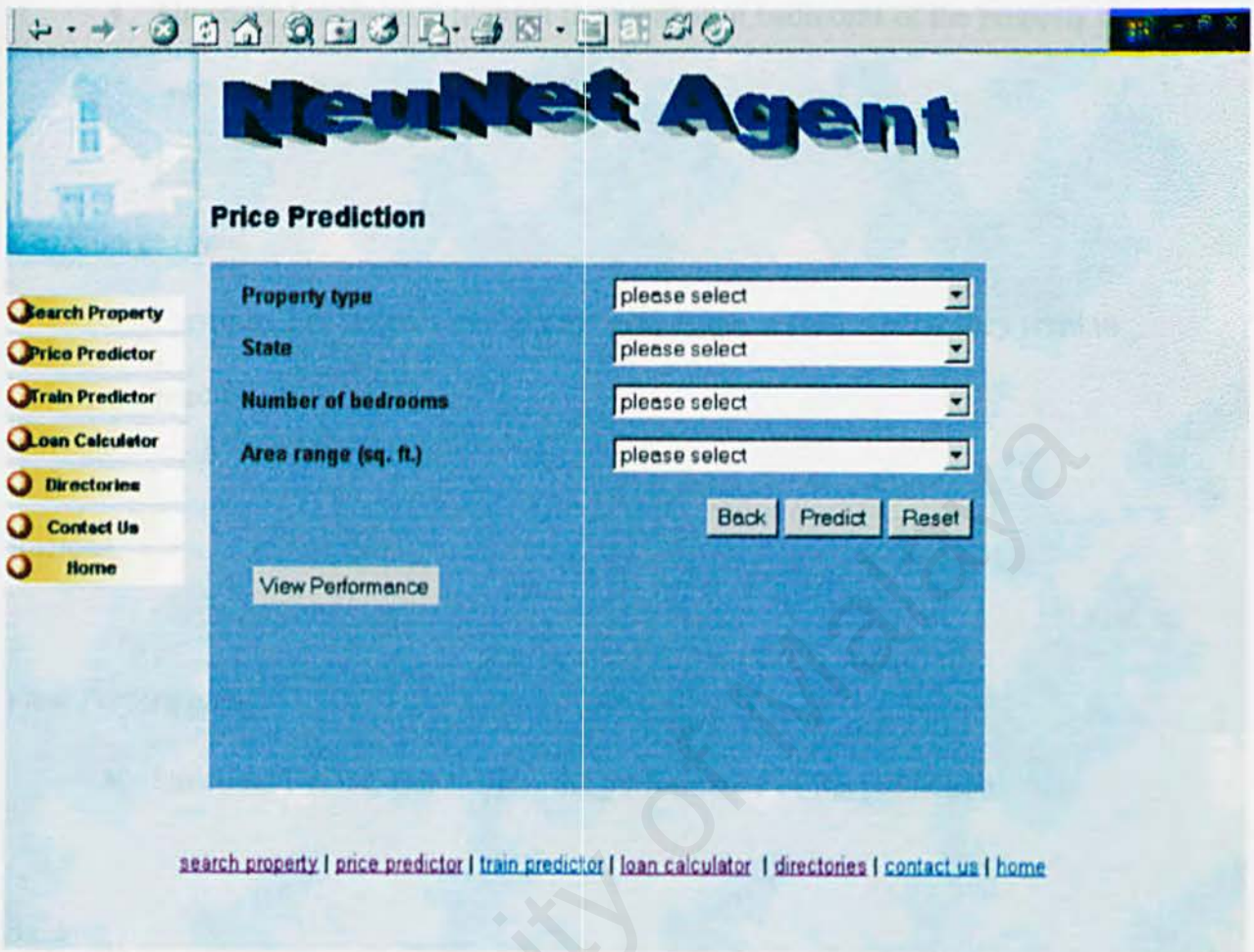
- Function: Let the user to submit their feedback form.

Reset

- Function: Reset all the fields in the contact us form.

University of Malaya

Price Predictor Page



Price Prediction Form

Property Type Field

- Function: Let the user to input the property type they want to predict.

State Field

- Function: Let the user to input the state of the property they want to predict.

Number of Bedroom Field

- Function: Let the user to input the number of bedrooms of the property they want to predict.

Area Range Field

- Function: Let the user to input the area range of the property they want to predict.

Buttons

View Performance

- Function: Let the user to view the performance of the prediction.

Back

- Function: Go to the previous page.

Submit

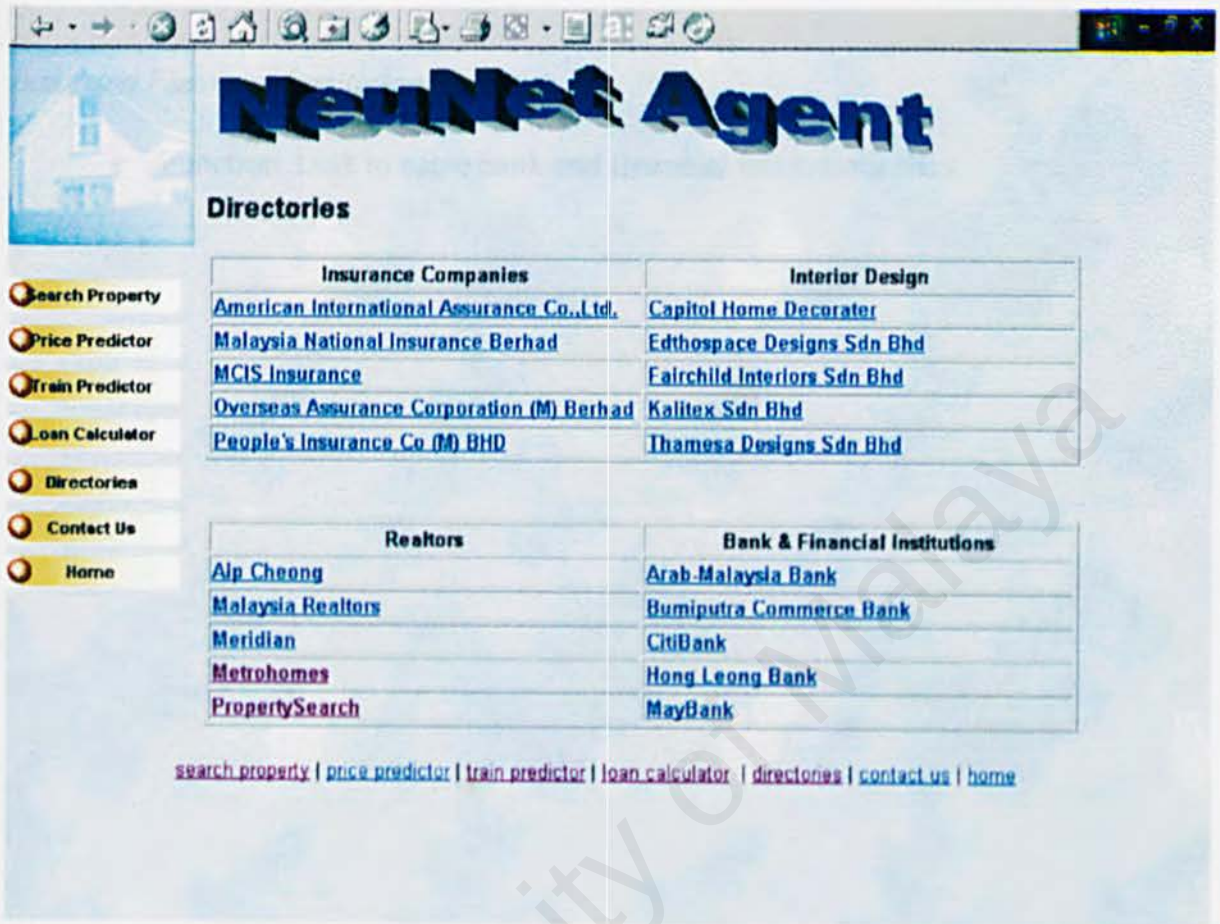
- Function: Let the user to submit their prediction form.

Reset

- Function: Reset all the fields in the prediction form.

University of Malaya

Directories



Links

Insurance Company

- Function: Link to some insurance company sites.

Interior Design

- Function: Link to some interior design company sites.

University of Malaya

Realtors

- Function: Link to some realtor's company sites.

Bank and Financial Institutions

- Function: Link to some bank and financial institutions sites.

University of Malaya