Palestine Polytechnic University

College of Administrative Science and Informatics

Information Technology department

Water Consumption Prediction in Hebron Using RBF Neural Networks

Bayan Ahmad Al-sarsour

Dima waleed Al-jabary

Hiba saleem Al-ju'beh

A final project submitted in partial fulfillment of the requirements for the degree of B.Sc. in Information Technology.

Supervisor

Dr. Mohammed Aldasht

2009

Abstract:

The lack of water in Palestine in general is a serious problem due to the geopolitical issues, as an urgent solution that will reduce the suffering of the citizen is to achieve a fair distribution for the available quantities of water; this requires a prediction of the water consumption for the Hebron citizen.

So, our work aims to predict the amount of water consumption for a given customer in a given season. The prediction is based on the customer's data during the last two years (2007, 2008) from the database of Hebron municipality.

In order to achieve the project goal, we implement a radial basis function network, to achieve acceptable prediction for future consumption of the water. Before using the input data, a normalization and processing of the data is carried out, to get the suitable input for the training phase of the neural network.

The results show that radial basis function networks are efficient when used to solve prediction problems. Also, the results show that the training method is the most important part when building the neural network. In the case of our project we noticed that the training still needs much work to reduce the error and to permit the network to accept the available input data with high variance.

Dedication:

To our parents and family ...

To our instructor Dr. Mohammed Aldasht...

To our friends ...

To the coming generation who will benefit from this project ...

To all martyrs who scarify them selves struggling toward freedom ...

And to all teachers, lectures...

Project team:
Bayan
Dima
Hiba

Acknowledgement:

First we would thank god for giving us bless of thinking and believing, we would like to acknowledge our gratitude to all those who have helped our in the writing of this project. First and for most we wish to express our sincere thanks to Dr. Mohammed Aldasht, and Dr. Hashim Tamimi for there insightful comments on an earlier version of the project. Finally we thank our parent for encouraging, and giving the suitable environments.

A.1.2 Nonfunctional requirements 38
A.2 Risks analysis 39
A.3 Time schedule and Gant chart 39
B: Implementation 41
B.1 Interface function 41
B.2 user interface design 44

Table of Contents:

Chapter One: Introduction	1
1.1 Overview	2
1.2 Problem definition	3
1.3 Related works	4
1.4 General Organization	5
Chapter Two: Theoretical Background	6
2.1 Overview	7
2.2 Regression and Prediction problems	7
2.3 Artificial Neural Networks	8
2.4 Radial Basis Function Neural Network	10
2.5 ADLINE training algorithm	13
2.6 Summery	14
Chapter Three: Methodology	15
3.1 Overview	16
3.2 RBF Neural Network	16
3.3 RBF Network Functions	19
3.4 Sequence diagram	21
3.5 Class diagram	22
3.6 Flowcharts	22
3.6.1 Normalization flowchart	22
3.6.2 RBF Parameters initialization flowchart	23
3.7 Summery	24
Chapter Four: Experimental Results	25
4.1 Overview	26
4.2 System developing environment	26
4.2.1 Hardware	26
4.3 Data Source	27
4.4 Experimental Results	28
Chapter Five: Conclusion and future work	32
5.1 Overview	33
5.2 Conclusion	33
5.3 Future work	33
References	34
Appendices	36
A: System planning & analysis:	36
A.1 System Requirements	37
A.1.1 Functional requirements	37

List of Tables:

Table 3.1: Classification of data for training and testing stages 17
Table 4.1: Hardware Requirement 26
Table 4.2: Testing Result 31
TableA.1: Time for each project task 39
Table A.2: Task and time schedule 40

List of Figures:

Figure 2.1 : Supervised learning & Unsupervised learning	10
Figure 2.2 :An RBF network with one output	11
Figure 2.3 :Gaussian radial basis function	12
Figure 3.1 : The 4 neural networks in each region	17
Figure 3.2 :Network architecture with input selection	18
Figure 3.3 :System sequence diagram	21
Figure 3.4 :Class Diagram	22
Figure 3.5 : Normalization Flowchart	23
Figure 3.6 :Initialization Flowchart	24
Figure 4.1 :Winter Training Error	28
Figure 4.2 : Spring Training Error	29
Figure 4.3 :Summer Training Error	29
Figure 4.4 :Autumn Training Error	29
Figure 4.5:winter prediction	30
Figure 4.6:spring prediction	30
Figure 4.7:summer prediction	30
Figure 4.8:Autumn prediction	31
Figure B.1: Navigation Flowchart	43
Figure B.2: Home page	44
Figure B.3: Project Page	44
Figure B.4: Prediction Process first step	45
Figure B.5: Prediction process second step	45
Figure B.6: Prediction process third step	46
Figure B.7: contact us	46

Chapter One: Introduction

1.1 Overview.
1.2 Problem definition.
1.3 Related works.

1.1 Overview

The lack of water in Palestine in general is a serious problem due to the geopolitical issues. As the rest of other Palestinian provinces, Hebron suffers from the lack of water due to the occupation policies and interventions to control the natural water resources, in addition to the controlling of the water resources to feed the settlements at the expense of the Palestinian citizen. Hebron is one of Palestinian cities which suffer from the scarcity of water resources and the increase in population.

This problem started drawing the attention of many people, so some conferences has been held such as First Annual Conference of the water in Hebron March 2009 . According to this conference, the main problem of water in Hebron returns to the small quantity that comes to the city. Hebron city had a population of more than 200 thousands people in 2008. The daily needed quantity for the city is more than 40000 cubic meters while an average of only 12000 cubic meters enters the city daily. According to Al-osaily declarations in the mentions conference "water shortage is a real crisis as the city of Hebron has a very limited number of water sources".

Other factors contribute to the problem are leakage and loss of water. In such a situation, a good distribution of the available quantities of water is essential to handle such a problem.

It is obvious that solutions to such a problem may include: seeking for sufficient water sources and offering the required quantities. Also, renewing the water distribution network and using the modern leak and loss detection tools are required to overcome such a problem.

An urgent solution that will reduce the suffering of the citizen is to achieve a fair distribution for the available quantities of water. This requires a prediction of the water consumption for the citizen. Such a prediction will be used to calculate the water quantities needed for each area in the city during a few next months. Also, the prediction will help in foreseeing any real water crisis before it happens. So, our project aims at providing a prediction model which will be used to achieve a fair distribution for the available quantities of water.

1.2 Problem definition

Hebron municipality has a large database for the water distribution system. This database contains a large history for the water meters. Such a history can be exploited to predict a future needs and distributions of water which in turns contributes partially in solving the water shortage problem in the Hebron city. Also, such predictions will aid in handling some problems like leak and loss detection.

The prediction methods are hard especially that of the future and the risk in the prediction may be unavoidable, before the process we didn't know the answer, but after the process we all know.

Learning prediction is very important in different scientific discipline, such as prediction of water distribution system, and there are several method to perform this task , there procedures use the information available about the behavior of the time series in the past, to predict the behavior in the future, therefore the prediction operation based on the consumer history data in the city of Hebron for the years of 2007 and 2008,because the diversity of the information is a critical factor that affect the prediction accuracy and realism .

This project aims at building a model for water consumption prediction. The model is based on RBFNN (Radial Basis Function Neural Network). Our approach is to use the history data of water consumption for the citizens in Hebron (which is divided into areas) to build four prediction networks for each area, one for each of the four seasons.

1.3 Related works

There are many works proposed in the literature to handle such a problem, most of these works aimed at dealing with leak and loss detection. Other works aimed at finding best distribution for some available water resources. Most of the proposed works were based on evolutionary algorithms.

Parma project presented in [1] aimed at developing an instrument to improve knowledge of the real operation of the system and to identify any critical areas in terms of water loss for reducing water usage, So through this method it is possible to determine the actual operation of water distribution network in a 24 hours period and to identify the areas with a large water loss. Parma implement a simulation model that is based on the genetic algorithm to improve the water distributions, and after several runs of the GA the minimization of the difference between pressure and flow measured in network and calculated by the model for an extended period simulation of 24 hours has allowed the identification of the most critical areas of the network in which most higher value of water losses is found.

In [2] present the result of the joining between the two projects, Water Software System (WSS) research group De Montfort University and Water operational Research Center (WORC) Burnel University in order to develop benchmarks to compare between algorithms that are used in water applications, and facilitate this comparison by providing a reference water network models as benchmarks, finally the result of the benchmarks helps to put the foundation for all applications with simulation as an example.

In [3] a study use neural network based method for simulation; an experiment was applied to water demand prediction in the yellow river basin. The objective of this study was to investigate the potential of using RBF network in water demand forecasting, first the process variables are obtained then a method is employed to train the values of the centers and widths of the RBF and update on the network weight is done, the study suggest water demand prediction process based on the RBF neural network, the number of hidden node were automatically determined, they use objective function(total or average error function) as an indication to change the structure of the neural network. Finally the result of the numerical simulation demonstrates the effectiveness of the previous method, thus the used method is not only suitable for water demand prediction but also suitable for forecasting process.

In [4] the direct and recursive prediction of time series are implemented and can be considered as modeling problem. The model is built between the input and the output, where the input is selected using the mutual information, and they used to predict the future value based on the previous value, and they use the first two thirds data of the whole dataset for training, and the remaining data for testing, and they found that the combination of the direct prediction and mutual information gives better performances than recursive prediction.

In [5] they deal with the problem of time series prediction from given set of input/output data, rearchers aim to access reliable prediction of future values based on past or present data, they use anew method of prediction of time series data using radial basis functions (RBFs) (real value function where the value depends on the distance for origin or distance from some point called center c).this approach based on anew efficient method of clustering of the centers of the RBF Neural Networks.

Researchers proposed an algorithm of clustering especially suited for function approximation problems. this method calculates the error committed in every cluster using the real output of the RBFNN ,this method does not only approximate the value of output ,but it also try to concrete more clusters in those input regions where the approximation error is bigger thus attempting to homogenize the contribution to the error of every cluster.

1.4 General Organization:

This work consist of five chapters, chapter one contain overview about the project, problem definition, and the related work ,and in chapter tow we will talk about theoretical background of the project such as regression and prediction problems, artificial neural network, and about radial basis function neural network, where chapter three explain our methodology, chapter four explain the experiment result that we achieved , finally, chapter five contain conclusion and future work.

Chapter Two: Theoretical Background

2.1 Overview
2.2 Regression and Prediction problems
2.3 Artificial Neural Networks
2.4 RBFNN
2.5 Summery

2.1 Overview

We predict the user consumption using neural network approach, which used to teaching the network .The history of the consumers consumption values in a selected limited time and applying the taught information to predict the consumption values in the future, in our project we will apply the supervised learning and build the learning function from training data which consist of input object and desired output.

In this chapter we will mention all the terms and topics that are needed to understand this project, So we will talk about the regression problem, and how it will be used to predict a value given the value of another variable, and about neural network approach, and its categories in term of their corresponding training algorithm, also we will briefly talk about radial basis function as the type of artificial neural network.

2.2 Regression and Prediction problems

In regression problems, the purpose of the neural network is to learn mapping from the input variables to continuous output variable. A network is successful at regression if it makes a more accurate prediction than simple estimate [7].

Linear regression is used to construct a simple formula that will predict a value or values for variable given the value of another variable. A regression model is often constructed based on certain conditions that must be verified for the model to fit the data well, and to be able to predict accurately. Hence, there is a confidence interval for a single future value of the output y corresponding to a chosen value of the input x.

In a simple linear regression, we predict scores on one variable from the score on a second variable. The variable we are predicting is called the *criterion variable* and is referred to as y. the variable we are basing our predictions on, is called the *predictor variable* and is referred to as x. When there is only one predictor variable, the prediction method is called simple regression [13].

2.3 Artificial Neural Networks

"An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. NN is an interconnected group of artificial neurons that uses a mathematical model or computational model for information processing based on a connectionist connection set approach to computation"[8], and the prediction of time series using neural network consists of teaching the network the history of the variable in a selected limited time, then applying the taught information to the future. Data from past are provided to the inputs of neural network and we expect data from future from the outputs of the network.

In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network, and it has many advantages such that its able to learn how to do tasks based on the data given for training or initial experience, also neural network can create its own organization or representation of the information it receives during learning time, and NN computations may be carried out in parallel, and special hardware devices are being designed and manufactured which take advantage of this capability, where the true power and advantage of neural networks lies in their ability to represent both linear and non-linear relationships and in their ability to learn these relationships directly from the data being modeled, and their ability to capture and represent complex input/output relationships.

"A key feature of neural networks is an iterative learning process in which data cases are presented to the network one at a time, and the weights associated with the input values are adjusted each time. After all cases are presented, the process often starts over again. During this learning phase, the network learns by adjusting the weights so as to be able to predict the correct class label of input samples. Neural network learning is also referred to as "connectionist learning," due to connections between the units. Advantages of neural networks include their high tolerance to noisy data, as well as their ability to classify patterns on which they have not been trained"[9].

The neural networks are commonly categorized in terms of their corresponding training algorithms, and there are two major learning paradigms, each corresponding to a particular abstract learning task. These are supervised learning, unsupervised learning [6].

Supervised learning: This learning method has been the mainstream of neural model development and it is a machine learning technique for learning a function from training data. The training data consist of pairs of input objects (typically vectors), and desired outputs, where the task of the supervised learner is to predict the value of the function for any valid input object, so that each output unit is told what its desired response to input signals ought to be. During the learning process global information may be required. Paradigms of supervised learning include error-correction learning, an important issue concerning supervised learning is the problem of error convergence which mean the minimization of error between the desired and computed unit values. The aim is to determine a set of weights which minimizes the error[14][15].

Unsupervised learning: In this technique, the training set consist of input training patterns only, and is based upon only local information. It is also referred to as self-organization, in the sense that it self-organizes data presented to the network and detects their emergent collective properties. Paradigms of unsupervised learning are competitive learning; with unsupervised learning it is possible to learn larger and more complex models than with supervised learning, this because in supervised learning one is trying to find the connection between two sets of observations. The difficulty of the learning task increases exponentially in the number of steps between the two sets and that is why supervised learning cannot[14][15].

In Figure 2.1 we can see that the supervised learning there is one set of observations called inputs, is assumed to be the cause of another set of observation which is called the output, but in unsupervised learning all the observations are assumed to be caused by a set of latent variables.

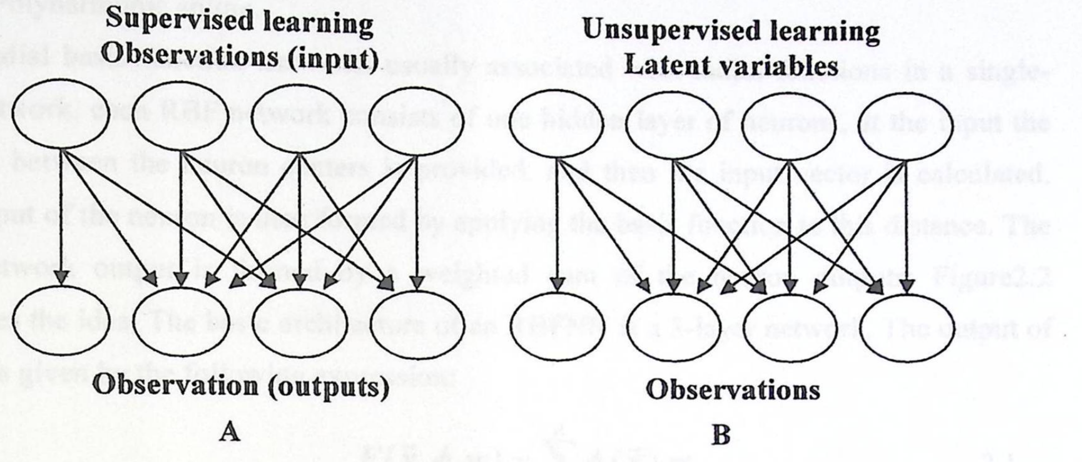


Figure 2.1: Supervised learning & Unsupervised learning adopted from [10]

"In unsupervised learning, all the observations are assumed to be caused by latent variables, that is, the observations are assumed to be at the end of the causal chain. In practice, models for supervised learning often leave the probability for inputs undefined. This model is not needed as long as the inputs are available, but if some of the input values are missing, it is not possible to infer anything about the outputs. If the inputs are also modelled, then missing inputs cause no problem since they can be considered latent variables as in unsupervised learning"[15].

2.4 Radial basis function neural network.

Radial basis function network is a type of artificial neural networks which is characterized by a transfer function in the hidden unit layer having radial symmetry with respect to a centre. RBFs use supervised learning for application to problems like: regression, classification and time series prediction. RBFNN were introduced into the neural network literature by Broomhead and Lowe which are motivated by observation on the local response in biologic neurons [11].

RBFNN has been extensively applied to wide range of fields such as forecasting, pattern recognition and signal processing. A radial basis function is a real value function; its value depends on its distance from some point called the center (origin).

There are a set of commonly used radial basis functions such as Gaussian, Multiquadric, Polyharmonic, Polyharmonic spline, thin plat spline which can be used as a special Polyharmonic spline.

Radial basis function networks usually associated with radial functions in a single-layer network, each RBF network consists of one hidden layer of neurons, at the input the distance between the neuron centers is provided, and then the input vector is calculated. The output of the neuron is then formed by applying the basis function to this distance. The RBF network output is formed by a weighted sum of the neuron outputs; Figure2.2 simplifies the idea. The basic architecture of an RBFNN is a 3-layer network. The output of the net is given by the following expression:

$$F(\vec{x},\phi,w)=\sum_{i=1}^{n}\phi_i(\vec{x}).w_i \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.1$$

Where $\Phi=\{\phi_i: i=1,\ldots, n\}$ are the basis functions set and w_i the associate weights for every RBF. The basis function ϕ can be calculated as a Gaussian function using the following expression:

$$\phi(\vec{x},\vec{c},r)=\exp(\frac{\|\vec{x}-\vec{c}\|}{r}) \quad \ldots\ldots\ldots\ldots\ldots\ldots 2.2$$

Where $\vec{c}$ is the central point of the function ϕ, r is its radius and $\vec{x}$ is the input vector.

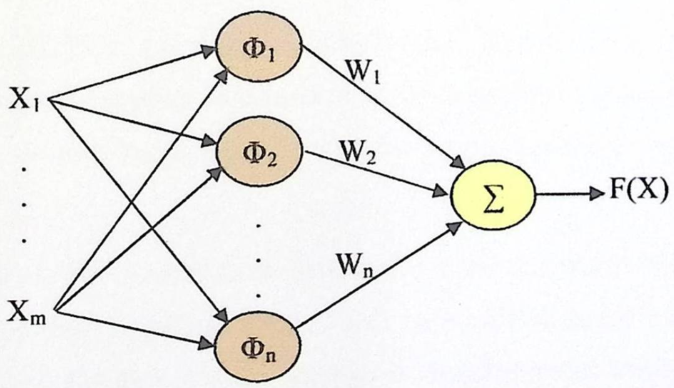


Figure 2.2: An RBF network with one output

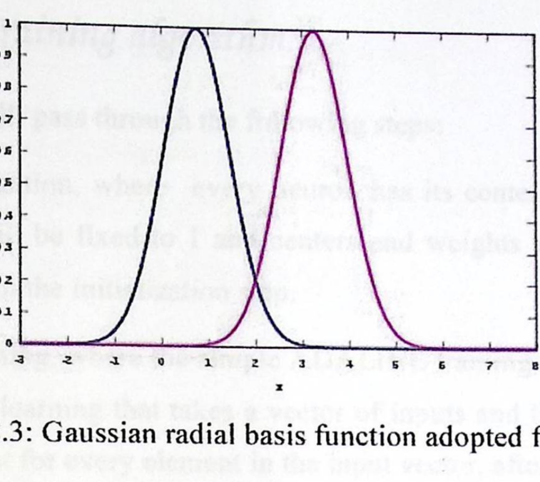


Figure 2.3: Gaussian radial basis function adopted from [12].

2.5 ADALINE training algorithm:

Building the network will pass through the following steps:

First step: Initialization, where every neuron has its center, radius, and weight. As stated before, radius will be fixed to 1 and centers and weights will be chosen randomly using random function in the initialization step.

Second step: Training, where the simple ADALINE training algorithm is applied, this is a type of supervised learning that takes a vector of inputs and the target from user, then the it initializes a weight for every element in the input vector, after that it calculates the net parameter depending on the weight and input value for all elements as follows:

$$net = net + w[i] * x[i] \quad \text{..............................}2.3$$

Where net is the result of the activation function, and w is the weight, and x is the input vector.

Then the algorithm will make some processing on the input vector to calculate output, then compare this output with the target, if the output equal the target the algorithm succeed else the weight will be updated and the net will be calculated again until reach the target or reach 20 iteration, in this situation the algorithm see that this is too much iterations and stop updating weight that's mean that the algorithm cannot access the target.

we apply ADALINE in our project as following : first we enter the input vector which contains the water consumption information in three seasons for all customers in an area, some processing on this input is made in order to predict the consumption in the forth season.

In our work we first initialize the parameter then after initialization the network will enter the training phase, in which weights will be modified based on the previous output and error, the process handles one training pair in each iteration, which is (x, d), where: x is the input and d is the desired output.

The following pseudocode shows the operation of training:

Repeat

Choose next training pair (x, d);

Compute network output at iteration i: $F_i(\vec{x},\phi,w)$

Compute error: $e(i) = d(i) - F_i(\vec{x},\phi,w)$

Update weights: $\mathbf{w}(i+1) = \mathbf{w}(i) + \eta e(i)\frac{F_i(\vec{x},\phi,w)}{w}$

Until weights converge to a steady set of values.

Where η is a small constant and is defined as the learning rate. The weight update method used here is called *the steepest (gradient) descent* method.

In each step, the difference between the output and the desired output is calculated as error, to check how much the output is close to the correct value. This algorithm is used to minimize the cost function $E(\mathbf{w}) = \sum_{i=1}^{n}\frac{1}{2}e^2(i)$ where *e (i)* is the error at iteration *i.* This is referred to as Least Mean Square (LMS).

2.6 Summery

In this chapter we talked about all the terms and topics that are needed to understand this project. So, we talked about the regression problem, and how it will be used to predict a value from another value, also we talked about neural network approach and its categories based on their training algorithm, and we talked about the radial basis function as the type of artificial neural network.

Chapter Three: Methodology

3.1 Overview
3.2 RBF Neural Network
3.3 RBF Network Functions
3.4 sequence diagram
3.5 class diagram
3.6 flowcharts
3.7 Summery

3.1 Overview

In our project, we will predict the water consumption for a consumer in a given season, and the prediction method is based on the consumer data during the last two years, then we will apply the methodology as presented in this chapter.

In this chapter we will mention all the terms, topics , and steps that are needed to apply water consumption prediction in successful and accurate way, So we will talk about the way to deal with the data taken from Hebron municipality, also we will take about classification concept and how we apply this concept in our project, finally we will explain how we build our network in sequential steps .

3.2 RBF Neural Network

This work aims to predict the amount of water consumption for a given customer in a given season. The prediction is based on the customer's data during the last two years from the database of Hebron municipality.

The available data represent the monthly consumption of each customer along the last two years. Customers are grouped into regions. We will need to build four networks for each region, a network per season, i.e., winter, spring, summer and autumn as shown in Figure 3.1. To make the learning depth in the number of customers we have classified the available data into two parts, the first will be used for training the networks and the second will be used for testing. The classification of the data is done in away that permits the representation of the different four seasons of the year for both training and testing without permitting the use of the same value in both phases, i.e., the value that is used in training phase will not be used in testing and vice versa. This classification of the data is shown in Table 3.1.

	Training				Testing			
	Winter	Spring	Summer	Autumn	Winter	Spring	Summer	Autumn
2007	Dec, Feb	Mar, May	June, Aug	Sep, Nov	Jan	April	July	Oct
2008	Jan	April	July	Oct	Dec, Feb	Mar, may	June, Aug	Sep, Nov

Table 3.1: Classification of data for training and testing stages.

As stated before, we need to construct four neural networks in each region, each one dictated to predict the amount of consumption in a given season, the neural network consisting of three layers, input, hidden, and output layer as shown in Figure 3.1.

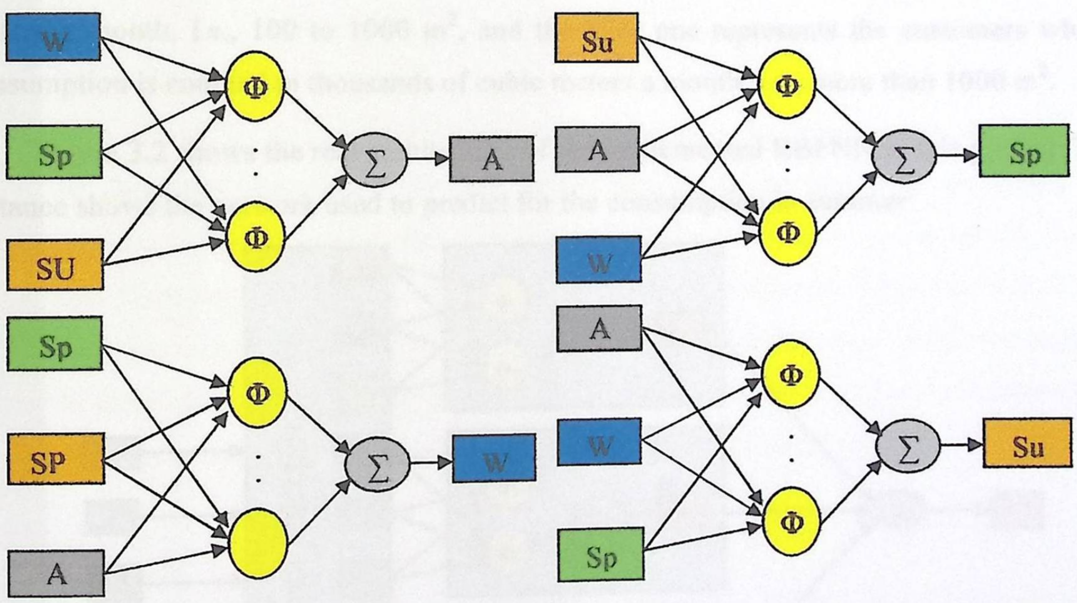


Figure 3.1: The 4 neural networks in each region. W: Winter, Sp: Spring, Su: Summer and A: Autumn

The data that we used as an input is the amount of water consumption of three seasons. The water consumption in a given season for a given customer is the average consumption of that customer in the corresponding three months of the season according the distribution given in table 1. e.g., the consumption in winter for a given customer is calculated as the average of his consumption in December and February, 2007 and January, 2008. The other inputs for the rest of seasons are calculated in a similar way. In fact, the calculated average is then normalized to the maximum to set between 0 and 1 to simplify the application of the Gaussian function, and then on the output side the predicted value is scaled to the original range to print the correct output. This normalization will imply the use of a radius equals to 1 for all the neurons in the network.

Since the water consumption is widely variable among different types of customers, i.e., home usage, agricultural usage, manufacturing usage, etc., an input selection is needed to classify the different types of customers so that they are grouped according to their consumption. We have decided to define three groups of consumers: the first group represents those whose consumption is counted in tens of cubic meters a month, i.e., 1 to 100, the second group represents those whose consumption is counted in hundreds of cubic meters a month, i.e., 100 to 1000 m^2, and the third one represents the customers whose consumption is counted in thousands of cubic meters a month, i.e., more than 1000 m^2.

Figure 3.2 shows the real architecture of the implemented RBFNN in this project. The instance shows the network used to predict for the consumption in summer:

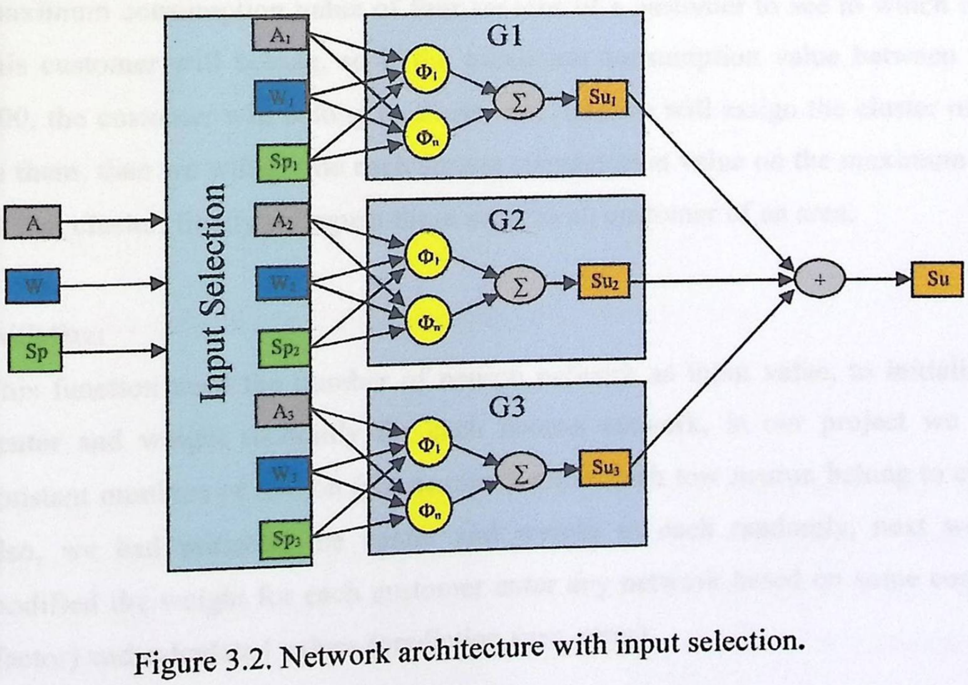


Figure 3.2. Network architecture with input selection.

3.3 RBF Network Functions

In this section we will take about our work function and operation. The application is divided into two main parts Neural Network & Interface, this part present RBF Neural network functions.

In NN part we develop a c++ program that can do too many functions each has its own input and output data as following:

- Load data:

  This function takes its input from an excel file which include the consumer consumption in four season in specific area, then save each consumer data to the input vector, finally an input vector which include all consumer .consumption data are returned.

- Normalization :

  The aim of this function is to limit the data values to be between zero and one, and to determine the cluster the user is belongs to. In the first step we found the maximum consumption value of four seasons of a customer to see to which cluster this customer will belong, so if the maximum consumption value between 0 and 100, the customer will belong to cluster one, and we will assign the cluster number to them, then we will divide each season consumption value on the maximum value of that cluster, finally we repeat these steps to all customer of an area.

- Initialize:

  This function need the number of neuron network as input value, to initialize the center and weight randomly for each neuron network, in our project we use a constant numbers of neuron (six neurons),where each tow neuron belong to cluster. also, we had initialize the center and weight to each randomly, next we will modified the weight for each customer enter any network based on some constants (factor) and calculated values (prediction sum, error).

- Training :

This function is the most important function in our project, we can say that this function represent the neural network, where the aim of this function is to predict consumption value based on customer input data, weight ,neuron centers, and radius , the aim of updating weight is to learn the network until reach acceptable prediction value. The number of iterations is determined, and utilizes alpha value (constant learning factor), then we entered the input value of the customer, and the center and radius to the first season.

The result is the prediction of the desired season, then we multiply the prediction value by weight, then the error value is calculated by subtracting the desired season value from the predicted value, this steps are repeated for all customers, and the average error value are calculated for all customers, then we entered the input value of customers to the second neuron. In addition, to the center and radius of this neuron, also the average error of the second neuron calculated, in the next iteration, we update the weight based on the error of the previous iteration of the same neuron each customer. If the previous error > 0, then apply the following equation:

$$W_{i+1} = w_i - \left(error * \frac{production\,value}{wi} \right) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.2$$

Else If the previous error<0 we will apply this equation

$$W_{i+1} = w_i - \left(error * \frac{production\,value}{wi} \right) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.3$$

Then the weight is updated until reaching an acceptable value.

- Testing:

The aim of this function is to calculate prediction value based on the best results that we achieved in the training phase, like the weight of each customer alpha value also the radius and center for each neuron. First we enter the customer input valuein addition to the last mention parameters, and then the predicted value is calculated.

3.4 sequence diagram

The following diagram in figure 3.3 represents the sequential steps of how prediction process implement.

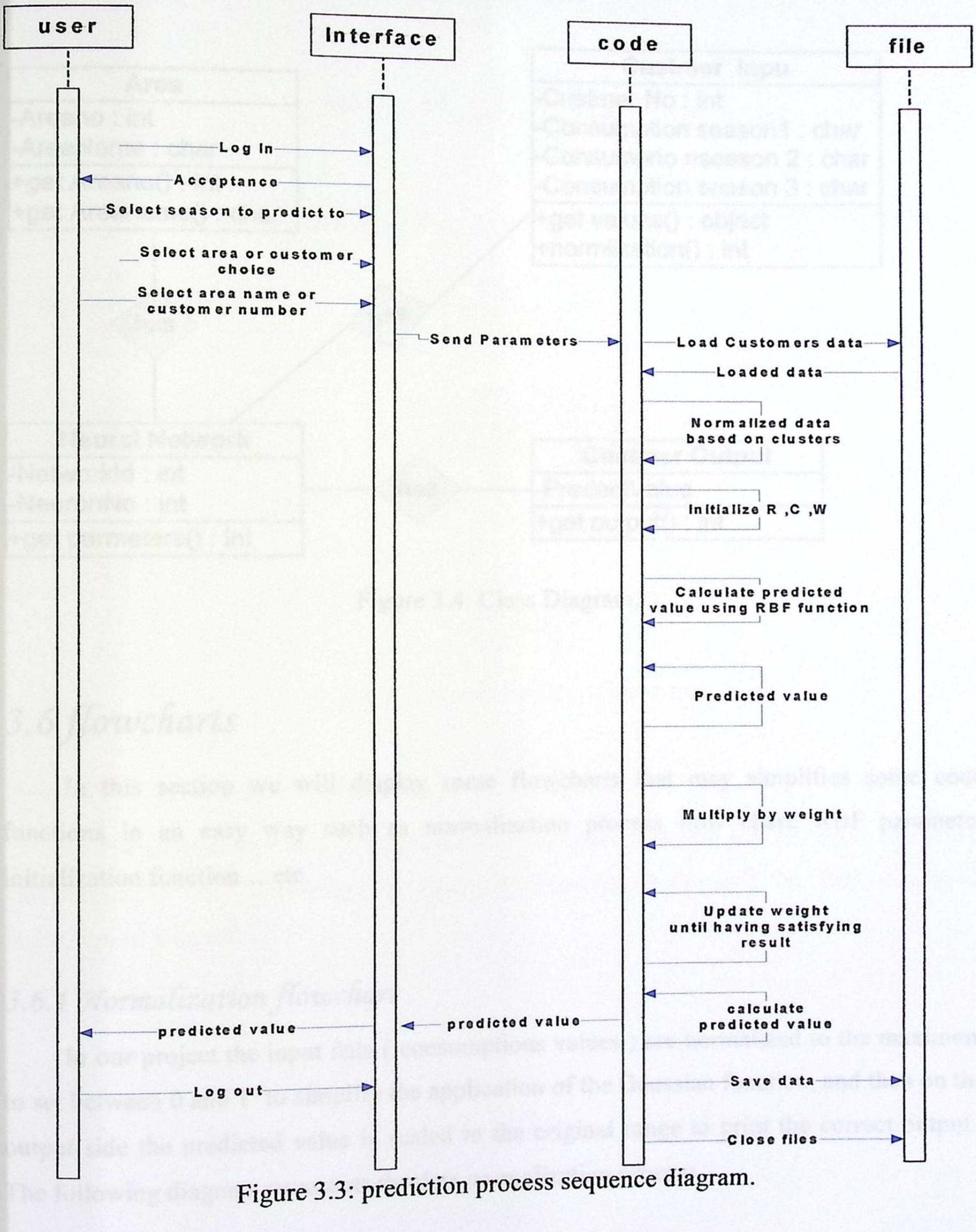


Figure 3.3: prediction process sequence diagram.

3.5 class diagram

Our project has many classes, and the following diagram in figure 3.4 represent all the classes, and the relationship between these classes.

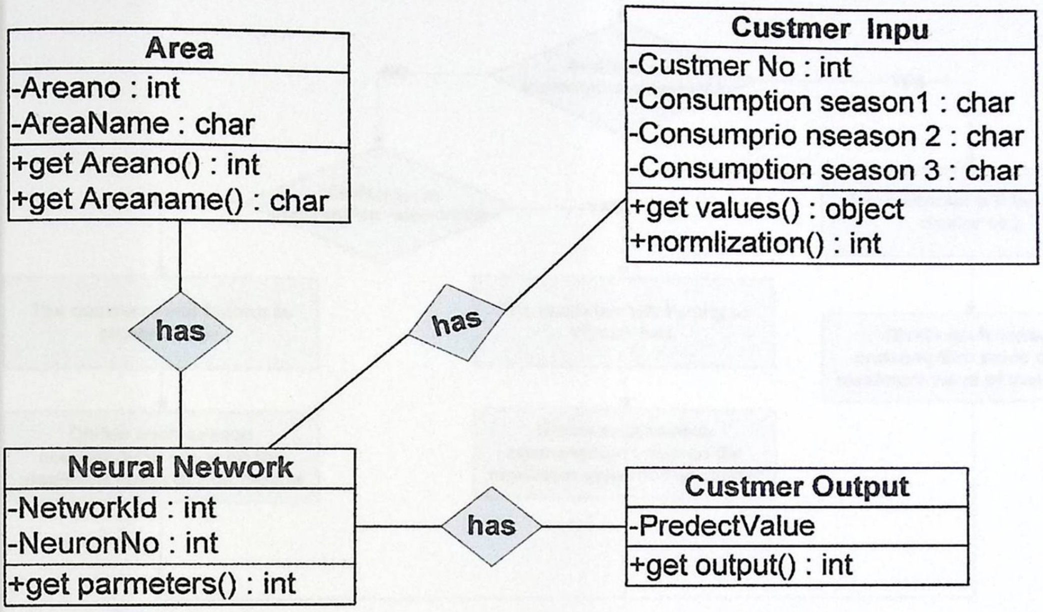


Figure 3.4: Class Diagram.

3.6 flowcharts

In this section we will display some flowcharts that may simplifies some code functions in an easy way such as normalization process flow chart, RBF parameter initialization function… etc.

3.6.1 Normalization flowchart

In our project the input data (consumptions values) are normalized to the maximum to set between 0 and 1 to simplify the application of the Gaussian function, and then on the output side the predicted value is scaled to the original range to print the correct output , The following diagram represents the data normalization process .

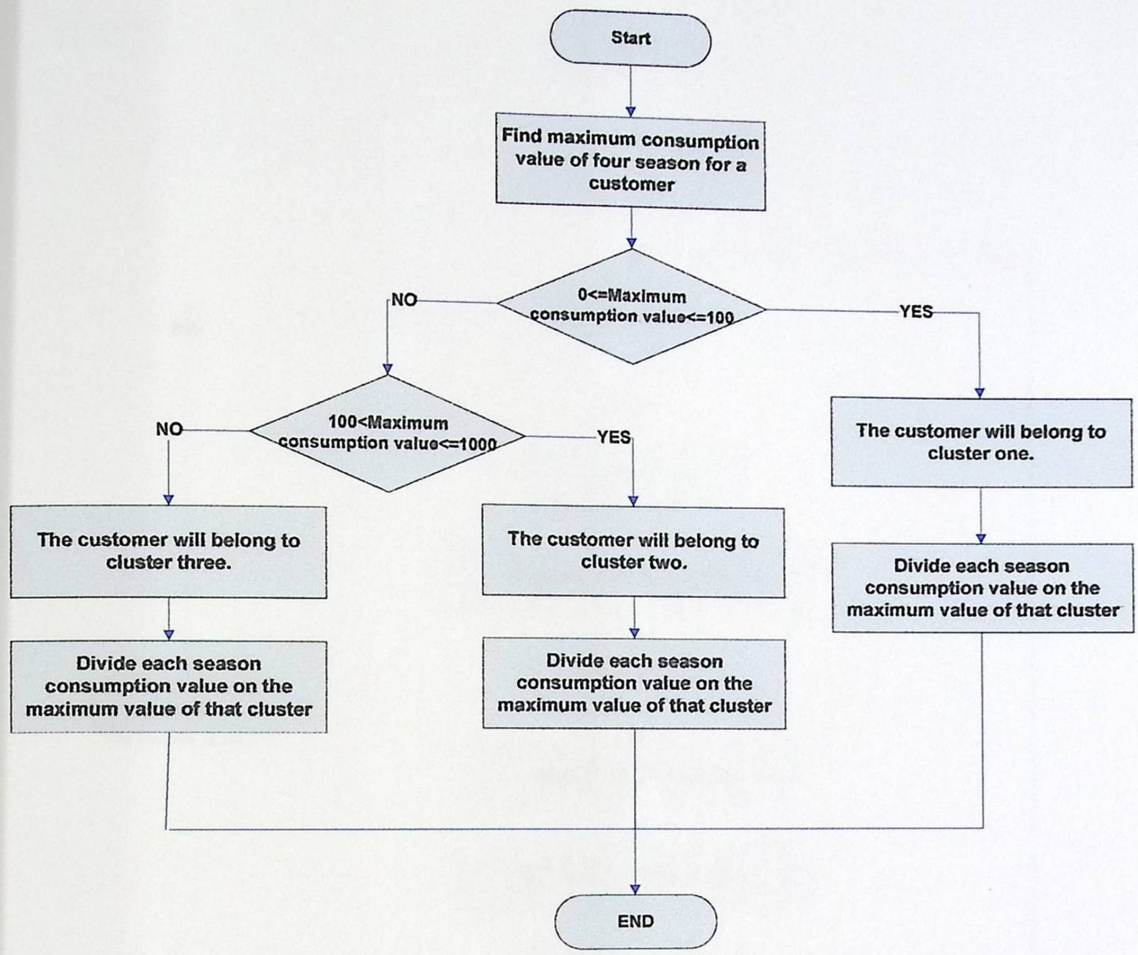


Figure 3.5: Normalization Flowchart

3.6.2 RBF Parameters initialization flowchart

In our project we use a constant numbers of neuron which we initialize the center and weight to each randomly; the following flowchart represents the RBF parameter initialization process.

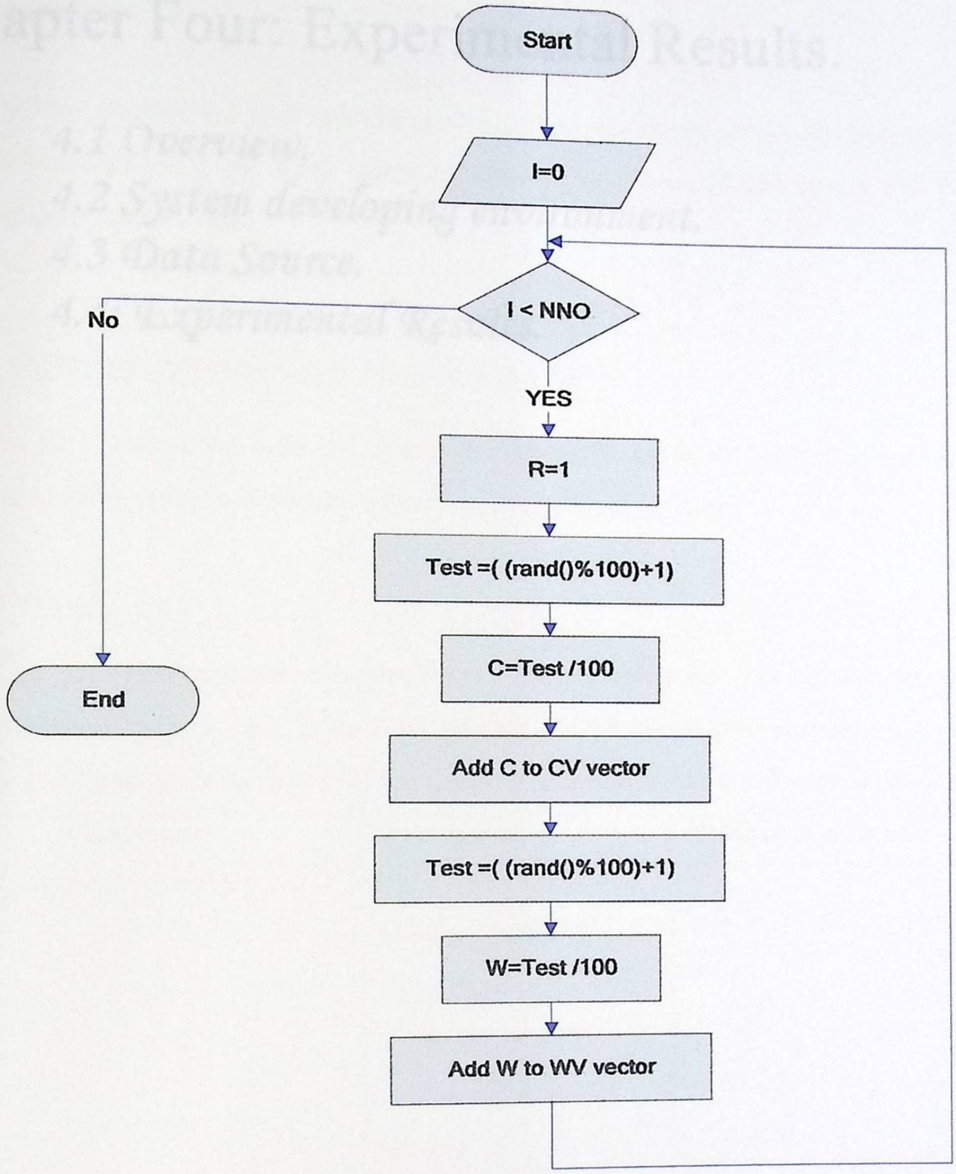


Figure 3.6: Initialization of RBF parameters .

3.7 Summery

In this chapter we talked about methodology, how we deal with the data, the main concepts such as classification are discussed, and some examples with some charts are given to simplify the idea. Also, we clarified deeply the main steps for building our network to make prediction process successfully.

Chapter Four: Experimental Results.

4.1 Overview.

4.2 System developing environment.

4.3 Data Source.

4.4: Experimental Results.

4.1 Overview

In this chapter we will talk about our project development environment also we will present conclusion of our work, some results will be presented and charts will present our work improvement steps.

4.2 System developing environment

In this section we will talk about the advances hardware and software requirement that we used in order to develop our project in professional way.

4.2.1 Hardware

We use three personal computers one for developing our project web site and second for programming, the third is used to prepare the project documentation, also we need a printer and flash memory ,our HW requirement characteristics are shown in table 4.1.

Component	Description	Number of devices
1. personal computer	CPU: 2400 MHz Memory: KB512 HD: GB40 CD ROM : 52X Modem: PCI 56 KB Mouse Keyboard	3
2. Printer	HP Laser Jet 1020 printer	1
3. Flash memory	1 G	3

Table 4.1: Hardware Requirement.

4.3 Data Source

Hebron municipality has a large database for the water distribution system, this database contains a large history for the water meters , which we exploited and used to predict the water consumption for a consumer in a given season , and the prediction method is based on the consumer data during the last two years (2007 and 2008) from the database of Hebron municipality, these input data are normalized first to limit all the values to be between 0 and one to simplify the application of the Gaussian function , then on the output side the predicted value is scaled to the original range to print the correct output .

4.4: Experimental Results

Results show that radial basis function networks are efficient when used to solve prediction problems. Also, results show that the training method is the most important part when building the neural network. In our project we predict the users consumption in four season with variant iteration and error rate. And These differences shown in figures below.

In Figure 4.1 we represent the iteration number and prediction error for the winter season.

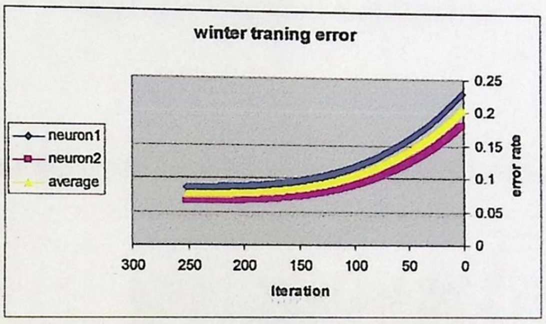


Figure 4.1: Winter Training Error

In figure 4.2 we represent the iteration number and prediction error for the spring .

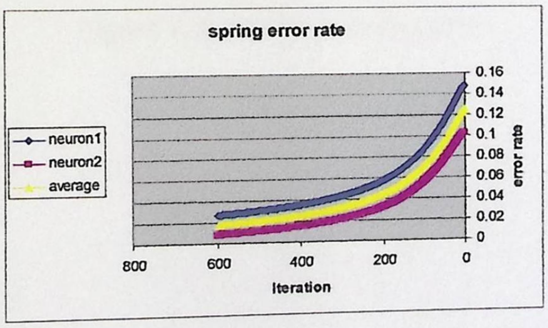


Figure 4.2: Spring Training error

In figure 4.3 we represent the iteration number and prediction error for the summer.

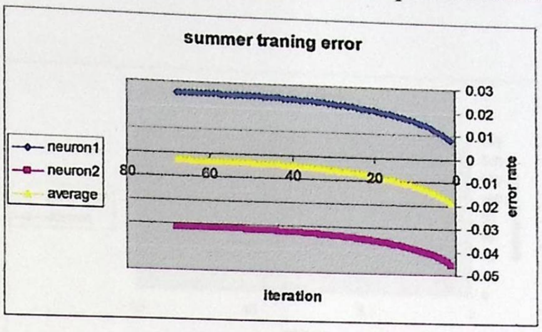


Figure 4.3 : summer training error.

In figure 4.3 we represent the iteration number and prediction error for the autumn .

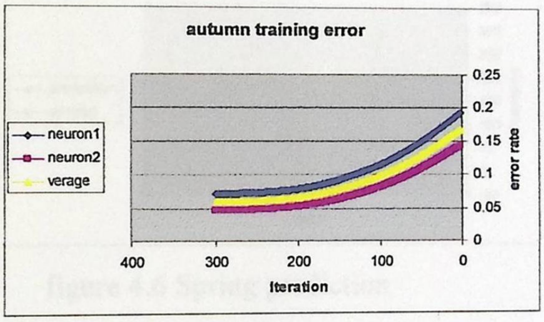


Figure 4.4: autumn training error.

The figures bellow show the relation between the desired and calculated prediction values in the four seasons:

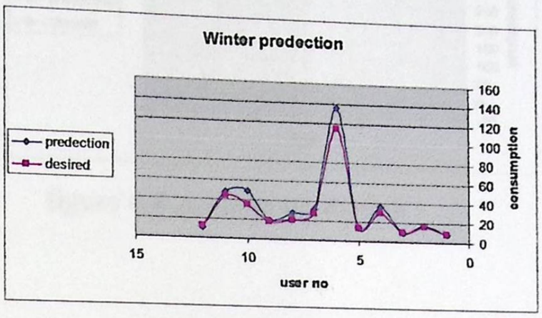


figure 4.5 winter prediction

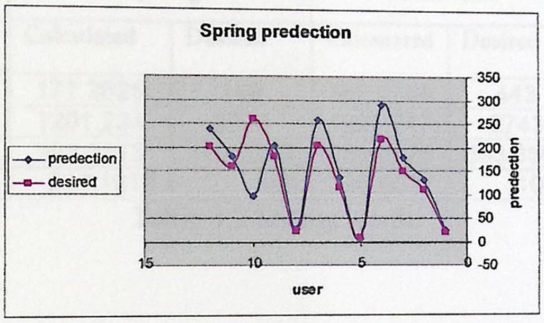


figure 4.6 Spring prediction

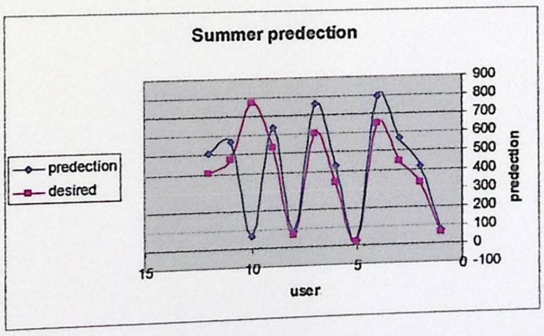


figure 4.7 Summer prediction

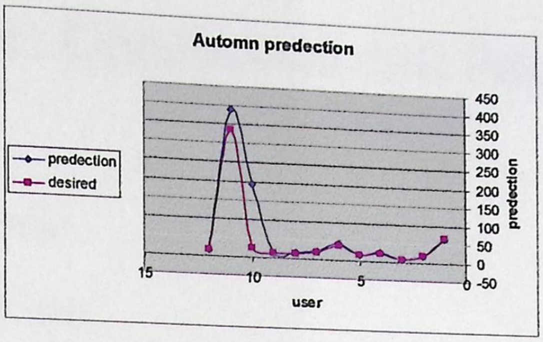


figure 4.8 Autumn prediction

The table bellows show a sample of user's predictions and desired values.

	Winter		Spring		Summer		Autumn	
Cno	Calculated	Desired	Calculated	Desired	Calculated	Desired	Calculated	Desired
2900	9.209892	9.666667	177.2825	149	567.0754	443	4.465444	5.666667
4207	39.21372	80.66666	1201.234	2743	1228.717	2743	78.50682	160.6667
5255	19.05933	17.55556	203.3313	179.2222	638.2526	533.6667	10.59887	10.66667
10622	29.27459	72	203.1616	503	396.6677	1004	1090.704	4.333333

Table 4.2 testing result.

Chapter Five: Conclusion and future work

5.1 Overview

5.2 Conclusion

5.3 Future work

5.1 Overview

In this chapter we will present a summary for our project work, also we will suggest some good future work which will make improvement on our work.

5.2 Conclusion

After the project is completed we can conclude the following:

1. A system is designed for water consumption prediction in Hebron city. the system employs for RBFNNs in each region to predict the consumption for the four seasons.
2. results show that RBFNNs are efficient to solve such a problem. Also, the convergence of the error in the training phase is shown to be very good and thus can give a precise prediction as shown in table 4.2.
3. The proposed system consists of two parts:
   a. RBF neural network based prediction: this part is used to achieve forecasting for the water consumption.
   b. An interface is provided to offer an easy interaction with the user.

5.3 Future work

As mentioned previously our project has faced many problem such as obtaining and processing the input data which have consumed much time. So, in order to solve the faced problems a series of future works should be done, the following are some suggested future works that will make a good improvement on our project:

1. Enhance the way of training to reduce the prediction error such enhancement could be achieved using advanced learning methods such as clustering techniques or evolutionary algorithms.
2. Use a dynamic architecture for the RBF network which permits changing the network structure during the learning phase.

References

[1] S.Bovis , E.Gerbino, "Genetic algorithms for water distribution network calibration: areal application" , Italy.

[2] Z. Yi Wu, "Water Loss Detection Via Genetic Algorithm Optimization-Based Model Calibration", 2006, Ohio.

[3] Y.wangs, J.Z.hang, "Water demand prediction based on RBF", 2008, china.

[4] Y.Ji, J.hao, N.Rayhani, "Direct and recursive prediction of time series using mutual information selection", Helsinhi university of technology, Finland.

[5] M. Awad, H. Pomares, I. Rojas, O. Salameh and M. Hamdon, "Prediction of Time series using RBF neural networks: a new Approach of Clustering" , Spain

[6] Neural network, <http://en.wikipedia.org/wiki/Neural_network>, accessed: 20 June 2009.

[7] Radial Basis Functions, <http://www.statsoft.com/textbook/glosr.html>, accessed: 20 June 2009.

[8] Prediction using neural networks,<http://www.obitko.com/tutorials/neural-network-prediction/neural-network-training.html>, accessed: 20 June 2009.

[9] Neural Networks Classification <http://www.resample.com/xlminer/help/NNC/NN Class_intro.htm>, accessed: 20 June 2009.

[10] H.Valpola, Supervised vs. unsupervised learning, 2000-10-31, <http://www.cis .hut.fi/harri/thesis/valpola_thesis/node34.html>, accessed: 20 June 2009.

[11] Center of cognitive science, university of Edinburgh, " Introduction to radial basis function networks", Scotland, April 1996.

[12] Unnormalized_radial_basis_functions <http://upload.wikimedia.org/wikipedia/ commons/thumb/f/fb/Unnormalized_radial_basis_functions.svg/350px-Unnormalized _radial_basis_functions.svg.png>, accessed: 20 June 2009.

[13] Introduction to Liner regression < http://onlinestatbook.com/chapter12/intro.html, accessed: 20 June 2009.

[14] Neural Network<http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html, Accessed:16 June2009

[15]supervisedVSunsupervised<http://www.cis.hut.fi/harri/thesis/valpola_thesis/node34.html, Accessed:21 June 2009.

Appendix

Appendix A: System planning & analysis
Appendix B: Interface Design and Implementation

Appendix A: System planning & analysis

A.1 System Requirements

In general, requirements are divided into functional requirements and non-functional requirements. Functional requirements are associated with specific task or function the system must support, while non-functional requirements are constraints on various attributes of these functions or tasks.

A.1.1 Functional requirements

Our project is divided into two main parts: neural network (NN) and Interface each has its own functional requirement.

We use c++ programming language to make a simulation for the neural network, so we design a program that predict customer or area water consumption in one of the year seasons depending on the previous customer behavior, after the predicted value is calculated our program compare it with the desired or real value and calculate the error rate, we aim to reach minimum error rate in order to have an accurate prediction.

In addition we design simple website using ASP.NET programming language in order to display our project results this web site have some main functional requirement as follows:

1. Predict water consumption value for customer or area in one of the year seasons.
2. Send some notes to the website administrator.
3. Display some charts about results.

A.1.2 Nonfunctional requirement:

As mentioned previously we have two main parts: neural network (NN) and Interface each has some nonfunctional requirement. The neural network part has some nonfunctional requirement:

1. Ability to produce results in short time.
2. Accuracy in calculation.
3. Efficiency in the parameter data type selection.

Our ASP.NET website provides user with the following nonfunctional requirement:

1. Accessibility to the website.
2. Security: make sure that user name and password are both correct.
3. Ease of use: our website is user-friendly interface, so it can be used easily.
4. Consistency: we use the same design for all the web site pages in order to guide the user.
5. Readability: background and font colors are comfortable for the user eyes.
6. Interactivity: user can interacted quickly with our website

A.2 Software environment:

We use many software programs while preparing our project such as:

1. Windows XP professional operating system.
2. Visual studio 2005: c++ language is used for design our neural network code part, while ASP.NET programming language is used to deign our project web site.
3. Photoshop CS: This software is used to design our web site background, buttons….
4. Microsoft Office 2003: We use Microsoft word used to prepare our project documentation; Microsoft Excel is used to make some reformat to the data taken from Hebron municipality, while Microsoft Power Point used to prepare our presentation.

5.Acrobat reader 9.1.: This SW used in order to read some previous studies and researches.

A.3Risk analysis:

A number of problems have been faced during the work; these problems can be summarized as follows:

1. Time: our project has limited time interval around 14 week only.
2. The Difficulty of data collection from Hebron municipality: data collection was very difficult step in the beginning of our project.
3. Scarcity of the previous studies and researches in the neural network field, and the lack of information resources for the RBF training methods.
4. Prediction error is still relatively large due to the high variance of the input data.
5. The idea that our project work on is relatively new.

A.4 Time schedule & Gantt chart

In this section we will talk about the expected time our project need to be complete in the desired way, also we will present the time schedule that show the main task and the actual time associated with each one).

The project tasks and the time needed for each one are explained in table A.1 below:

Task number	Task name	Time
1	Planning	1 week
2	Data collection	2 weeks
3	Analyze	3 weeks
4.	Design	2 weeks
5.	Coding	7 weeks
6.	Testing	2week
7.	Maintenance	2 week
8.	Documentation	14 week

Table A.1: Time for each project task

Gantt chart:

Table A. 2 Shows our project main task with the time associated with each one.

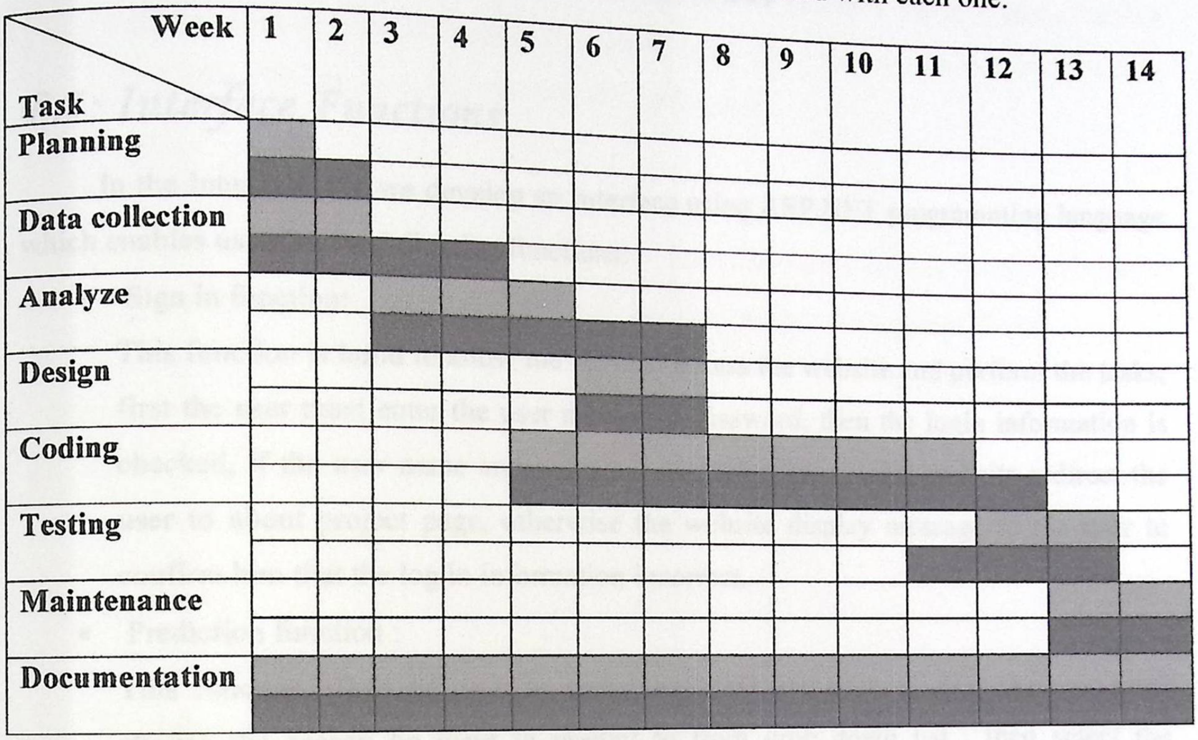


Table A.2: Task and time schedule.

Appendix B: Interface Design and Implementation

B.1: Interface Functions

In the Interface part we develop an interface using ASP.NET programming language which enables us to do the following functions:

- Sign in function:

  This function is build to allow the users to access the website and perform the tasks, first the user must enter the user name and password, then the login information is checked, if the user name and password entered correctly, the website redirect the user to about project page, otherwise the website display message to the user to confirm him that the log in information incorrect.

- Prediction function :

  This function allow the user to make the prediction process ,first the user must choose the season he want to predict to from drop down list , then select the prediction type (either consumer or area) , if the user select area choice , then the website will display dropdown list which include different areas name , then the user choose the name of the area he want to predict to , but if the user select consumer choice from drop down list , the consumers number will displayed in drop dawn list , then the user select the consumer he want to make consumption prediction to, the predicted value for either area or consumer will appear in label .

- Draw chart function :

  This function connect our website with the Google chart service, which help us to draw the wanted chart, first the user must choose the chart name he want to draw from dropdown list, then select the chart type (either pie chart or line chart), and select the chart size, where these parameter are sent to the Google chart services which return the wanted chart to the chart page.

- Make tower function :

This is learning function , which aim to help the user to know more about the website , and let the user know how to use the web function , there is many leaning choice such as learn by video, flowchart , or by text , the user must select the way he prefer, the learning information will be displayed .

After we complete our project functionality we find that we need a user friendly interface to present our work, so we develop user friendly interface using ASP.NET programming language, in this section we will describe our interface design using:

1) Navigation flowchart: Where it is to clarify the relations of the system screens with each other and their integration.
2) User interface design: where it is to display the design of the input and output screens, which reflect the interaction between the user and the system.

- *Navigation flowchart*

The following flowchart represents the general usage way of water consumption prediction website.

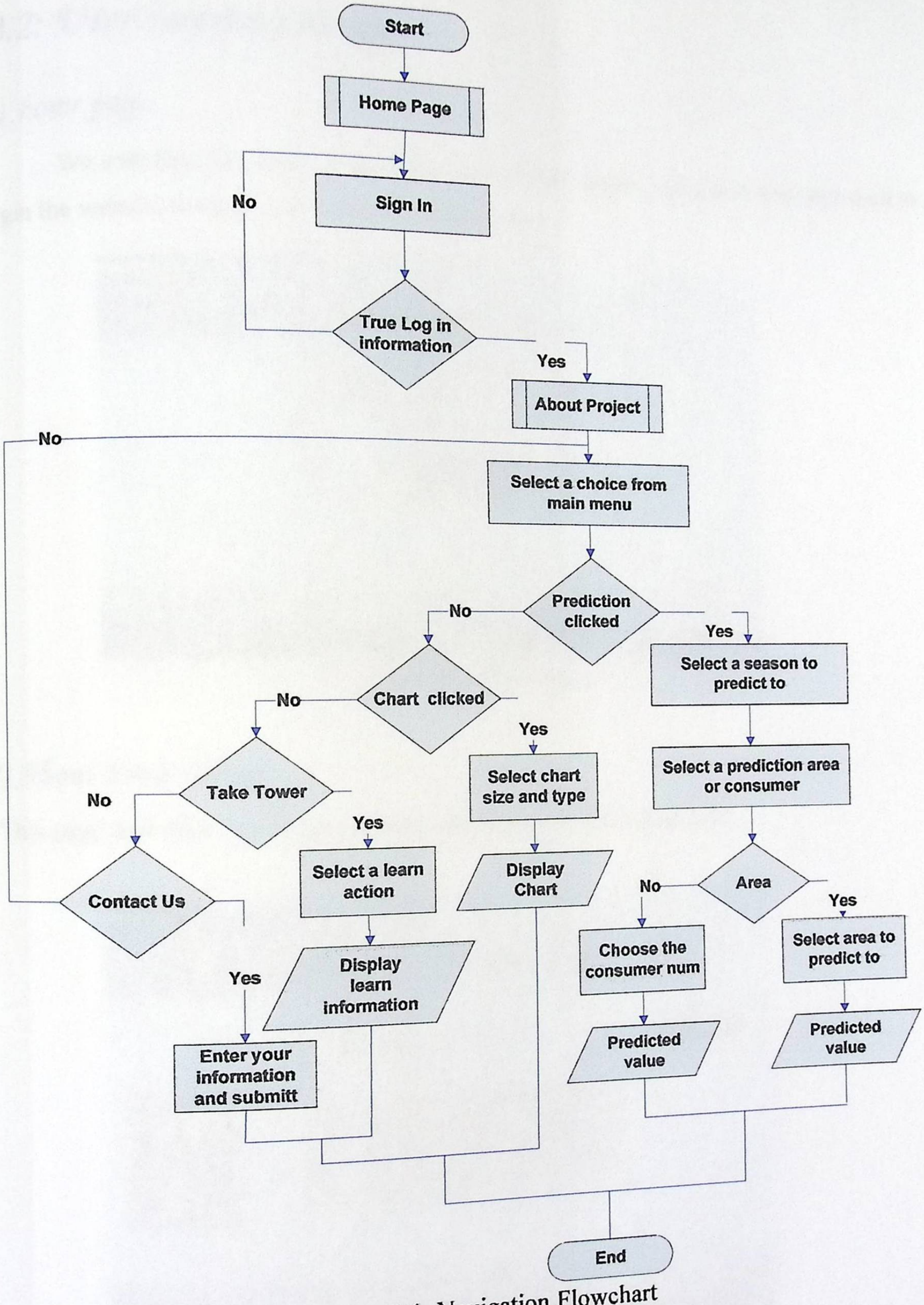


Figure B.1: Navigation Flowchart

B.2: User interface design

1) home page:

We consider this page as the main page of the website, since that this page used to login the website through user name and password.

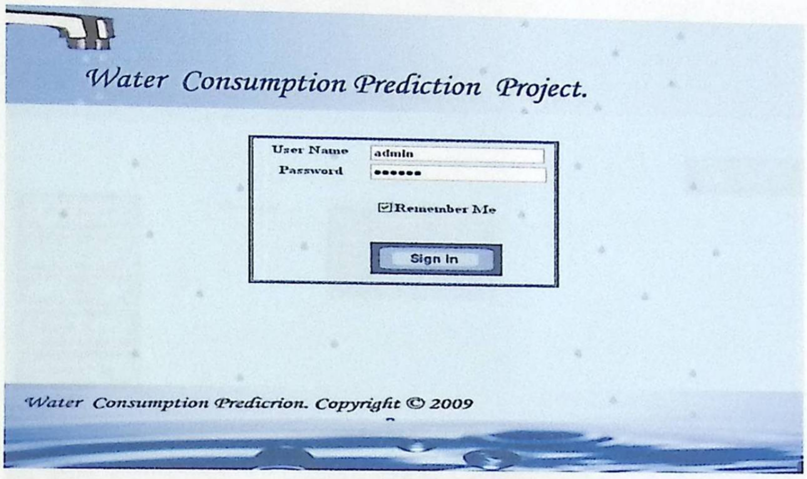


Figure B.2 :home page

2) About Project Page:

This page describes our project briefly, and talk about the project goal.

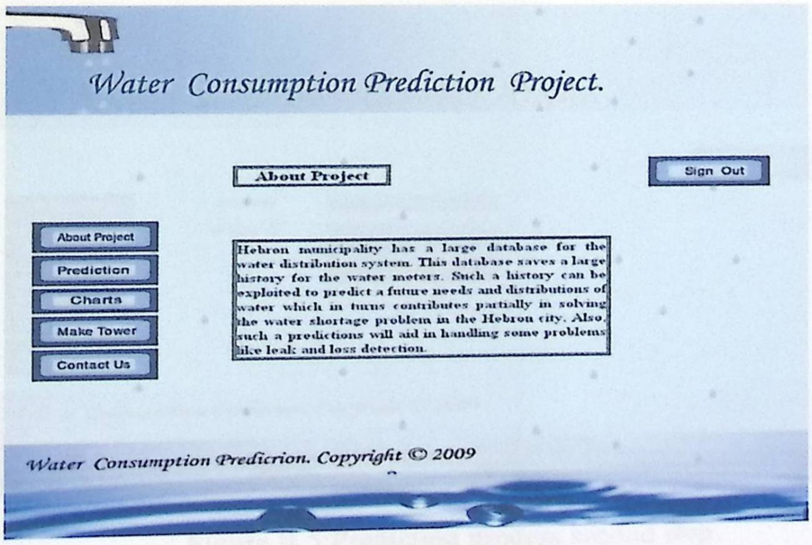


Figure B.3: About Project Page.

3) Prediction process

The core function of this page is to make prediction process for the year 2009, where the prediction can be for the specific user or area in specific season.

- First user must choose the season from dropdown list to predict to.

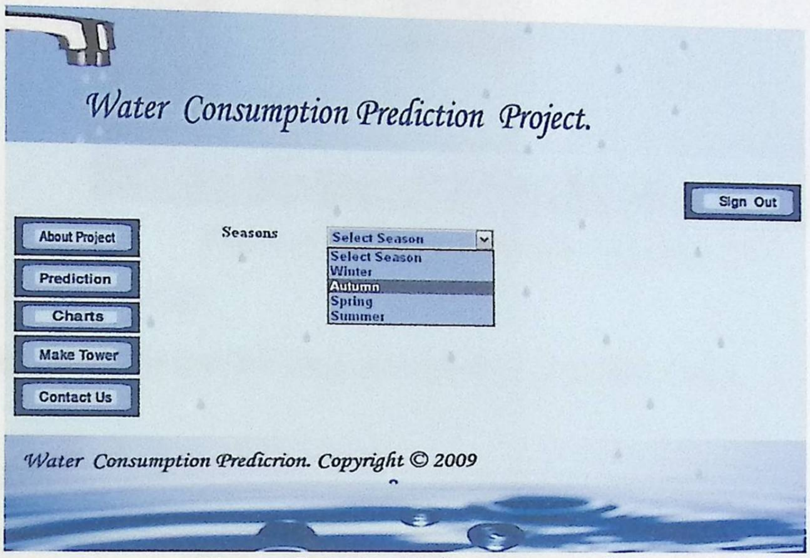


Figure B.4: Prediction Process first step

- Second, the user must choose the prediction type (customer or area) from dropdown list.

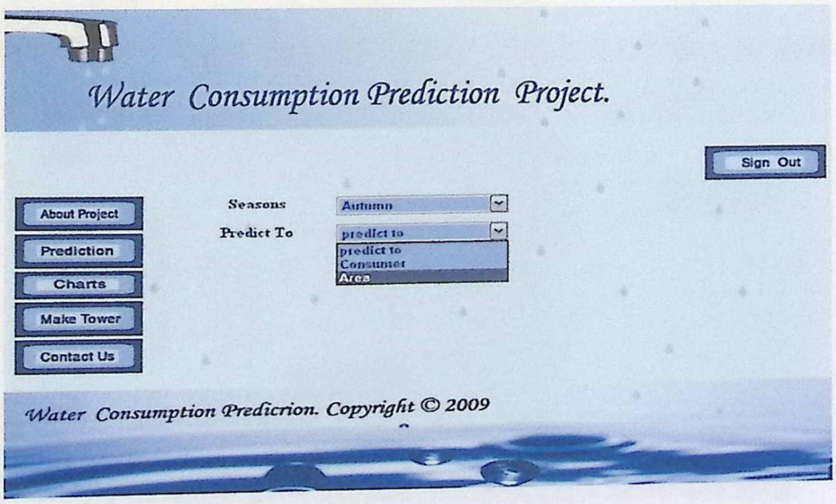


Figure B.5:Prediction process second step.

- Then, the water consumption data for the selected type are loaded from an excel file, then the predicted value will be displayed in label.

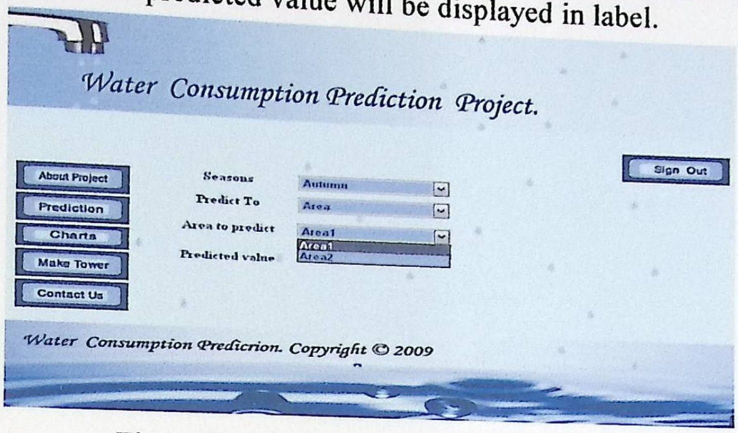


Figure B.6:prediction process third step.

4) Make a Tower page

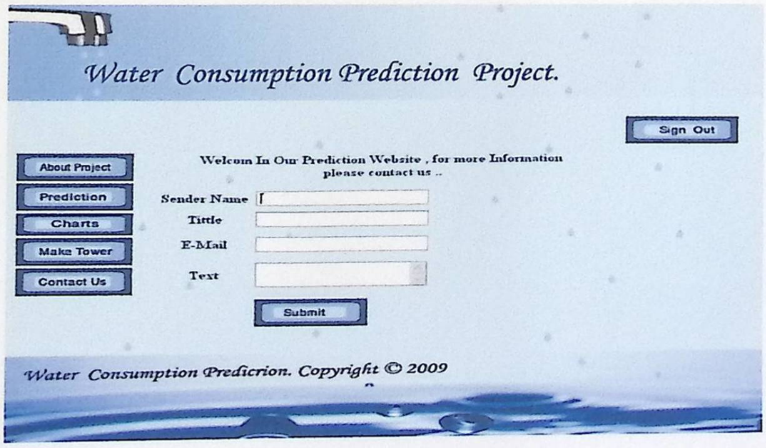
This page guides the user for using the website in successive way.

5) Chart page

In this page we connect our website with the internet Google services, where the chart parameters are taken from notepad, then the parameters and chart size and type send to the Google chart services.

6) Contact us page.

In this page the user can contact immediately with administrator, by sending letter.



B.7: contact us.