Chapter 1
INTRODUCTION AND SURVEY OF LITERATURE

1.1 General Introduction

In Bioinformatics, Information Technology is applied to the management and analysis of biological data about genes, proteins, nucleotides, amino acids pertaining to human beings and other living organisms. Although the terms bioinformatics and computational biology are often used interchangeably, bioinformatics more properly refers to the creation and advancement of algorithms, computational and statistical techniques to solve formal and practical problems inspired from the management and analysis of biological data. Computational biology, on the other hand, refers to hypothesis-driven investigation of a specific biological problem using computers, carried out with experimental or simulated data, with the primary goal of discovery and the advancement of biological knowledge. Bioinformatics is concerned with the information while computational biology is concerned with the hypotheses. Bioinformatics is the science of managing, analyzing, extracting, and interpreting information from biological sequences and molecules. Bioinformatics is a field that meets the needs of Biologists to process and interpret large volumes of data, obtained in genomics, post-genomics and proteomics.

Over the last few decades genomic biological data of various living organisms has been stored in large databases. Computational techniques have been developed to compare and analyze the biological information collected from different organisms.

Microarray data are widely used in genomic research due to the potential in gene expression profiling, facilitating the prognosis and the discovering of subtypes of diseases. Microarrays are the emerging technologies provide novel insights into gene expression and gene regulation.
1.2 Fundamentals of a Microarray

Microarray technology makes use of the sequence resources created by current genome projects and other sequencing efforts to identify the genes (Brown and Botstein 1999). Measuring gene expression levels in variable conditions provides biologists with a better understanding of gene functions, and has wide applications in life sciences. For example, microarrays allow comparison of gene expression between normal and cancerous cells. The technology has been referred by various names: DNA microarrays, DNA arrays, DNA chips, and gene chips. A Microarray is a glass slide or a small chip, onto which tens of thousands of DNA molecules are attached at fixed locations, which are called spots, each relates to a single gene (Bittner et al. 2000). A microarray experiment begins with good experimental design. After carrying out the biological experiment, the samples, either tissues from patient or animal model, or cells from in vitro cultures, are collected. Two samples of mRNAs are reverse transcribed into cDNA and are labeled with different fluorescent dyes, and co-hybridized onto a microarray. The hybridized microarrays are scanned and are imaged to acquire the fluorescence intensities for each spot on the glass slide, and the principle of microarray is as shown in the Fig.1.2.1

![Fig.1.2.1. The microarray principle. Image taken from http://www.transcriptome.ens.fr/sgdb/presentation/principle.php](http://www.transcriptome.ens.fr/sgdb/presentation/principle.php)

Image analysis is performed to obtain the raw signal data for every spot (Draghici, 2003; Schena 2002; Jung and Cho, 2002). The process of extracting features for further analysis from a microarray image can be categorized into three main steps: gridding, segmentation and quantification. Gridding is to assign each spot with individual coordinates. Segmentation is to classify the pixels into foreground, background and others (Wu and Yan, 2003). Quantification is to compute unique intensity values for each spot, which are related to the quantity of mRNA present in the solution that hybridize at the particular location of a microarray substrate. The data extracted from image analysis need to be pre-processed to
exclude poor-quality data and high quality data are normalized. The steps to convert raw image data into gene expression matrix is as shown in Fig. 1.2.2

![Diagram showing steps to convert raw image data into gene expression matrix](http://www.ebi.ac.uk/microarray/biology_intro.html)

**Fig. 1.2.2. Steps to convert raw image data into a gene expression matrix (from the EBI microarray site, http://www.ebi.ac.uk/microarray/biology_intro.html)**

**Types of studies those can be conducted with microarrays**

There are three types of applications of DNA microarrays (Tarca et al., 2006; Liang and Kachalo 2002).

1. Finding the genes that are differentially expressed across two kinds of samples or tissues acquired in two different experimental conditions. Differentially expressed genes are genes which have significantly different expression in two defined groups of microarray experiments. This is called a “class comparison” experiment (e.g., identification of genes differentially expressed in the placenta from normal pregnant women and women with pre-eclampsia).

2. Identifying the class membership of a sample based on its gene expression is known as class prediction. Class prediction is the process of construction of a classifier that analyzes the gene expression profile of a sample and predicts its class membership. The classifier is constructed based on a set of training samples and validated with samples from known class membership.

3. Discovering groups of genes that share same biological functionality i.e., Co-expressed genes, which requires clustering analysis and is known as class discovery. A summary of all the fundamentals are visualized step-by-step in the Fig. 1.2.3
1.3 Focus of Study

The focus of the proposed study is to discuss, formulate new algorithms and implementation of them on public and synthetic data as discovery of groups is a vital step in the analysis of microarray data and details are mentioned in the following paragraphs.

Effective Identification of co-expressed genes and coherent expression patterns in gene expression microarray data is an important data analysis task in bioinformatics, genomics research and biomedical applications. Microarray technology can simultaneously monitor the expression levels of thousands of genes during important biological processes and across collections of related samples (Schena et al. 1995, Schena 2000). There are two reasons for interest in co-expressed genes. First, there is evidence that many functionally related genes are co-expressed (Eisen et al. 1998; Spellman et al. 1998). For example, genes coding for elements of a protein complex are likely to have similar expression patterns. Hence, grouping Open Reading Frames (ORFs) with similar expression levels can
reveal the function of previously uncharacterized genes. The second reason for interest in co-expressed genes is that co-expression may reveal much about the genes’ regulatory systems. For example, if a single regulatory system controls two genes, then we might expect the genes to be co-expressed. In general, there is likely to be a relationship between co-expression and co-regulation.

Data clustering methods were proved to be a successful data mining technique in the analysis of gene expression data. Clustering is an unsupervised classification technique that aims at grouping a set of unlabeled objects into meaningful clusters (Jain and Dubes 1988, Grabmeier and Rudolph 2002) with the requirement that the resulting groups are homogeneous (that is, pairs of objects in the same cluster are highly similar) and neatly separated (that is, objects within distinct clusters are very dissimilar). In gene expression analysis, clustering groups the genes into biologically relevant clusters with similar expression patterns so that the genes clustered together tend to be functionally related. Clustering can reveal the co-expression of genes which were uncharacterized / unnoticed / unexpected previously which are having common characteristics or functional features. This has an important implication in pharmaceutical and clinical research. For example, by comparing gene expression data in normal and diseased cells, disease genes can be identified to facilitate drug design and therapy planning. Each cluster is considered as a group of co-expressed genes and the coherent expression pattern can be simply be represented by the centroid of the gene expression profiles in that cluster. Iyer et al. recorded the expression profiles of 517 human genes in twelve point time-series and gave a list of ten groups of co-expressed genes and the corresponding coherent patterns (Iyer et al. 1999).

1.4 Challenges in Clustering Process

Many clustering algorithms are proposed and developed in the literature to cluster the gene expression data. Hierarchical clustering has been applied to identify groups of co-expressed genes in yeast data set (Eisen et al. 1998). A two-way clustering technique (Alon et al. 1999) used to detect groups of correlated genes and tissues. Tamayo et al. (1999) applied self-organizing maps to identify clusters in the yeast cell cycle data set and human hematopoietic differentiation data set. Despite some successes with existing clustering algorithms in gene expression data analysis (Kohonen, 1989; Eisen et al., 1998; Ben-Dor et
al. 1999; Alon et al. 1999; Shamir and Sharan 2000; Smet et al. 2002; Zheng et al. 2003; Jiang et al. 2004), there is still no single clustering algorithm that is the most “dominant” gene expression microarray data clustering algorithm. A number of clustering methods have been proposed for gene expression data (Carpenter and Grossberg 1987; Chen et al. 1996; Guha et al. 1998, 1999; Kohonen et al. 1990; Zhang et al. 1996). k-means is the most popular partitional clustering technique for its efficiency and simplicity in clustering large data sets (Forgy 1965; Macqueen 1967; Hartigan and Wong 1979; Lloyd 1982, Jain 2010). The k-means was voted as one of the top ten algorithms in data mining (Wu et al. 2008). There are two major issues in the application of partitional based (k-Means-type) algorithms in cluster analysis.

- Number of clusters is to be specified in advance
- Sensitive to the initial centroids or seeds
- Produce different results in different independent runs

1.5 Problem Statement

The problem addressed in this study is to develop new clustering algorithms to
- identify optimal centroids for k-means type algorithms,
- select optimal values for the parameters to be supplied in advance,
- to determine the optimal clusters automatically and,
- to produce single optimal clustering solution automatically for the given data set.

Single Optimal centroids for k-means type algorithms are selected by choosing highest density point, which is close to more number of other points in the data set. Optimal clusters are identified by merging the clusters after validating with a cluster validity measure. The experimental results of the algorithms are shown that the proposed methods could find optimal clusters in terms of compactness and separation and also proved that these can group co-expressed genes in a better way.

1.6 Survey of Literature

DNA microarrays are emerging technologies in Bioinformatics, which are used to analyze the expression levels of DNA, and have many applications in Pharmacology, Medical diagnosis, Environmental engineering and Biological sciences (Schena et al. 1995). A microarray is a biological assay, with an ordered array of nucleic acids, proteins, or small molecules that enables the parallel analyses of complex biochemical samples
(Schena et al., 1995). Rapid advances in microarray technologies over the last several years have made it possible to monitor the expression of thousands of genes in parallel over many experimental conditions (Lockhart and Winzeler 1996). Large amounts of gene expression data have been generated by researchers (DeRaisi et al., 1997, Wen et al., 1998).

Clustering is a useful exploratory technique for analysis of gene expression data. Many clustering algorithms have been proposed for gene expression data. Three popular clustering algorithms have been used to cluster gene expression microarray data. They include the hierarchical clustering algorithm (Ward 1963), the k-means algorithm (MacQueen 1967) and the Self Organizing Map (SOM) (Kohonen 1990). Chen et al., (1996) used clustering to identify genes that have similar expression patterns to reduce the size of the regulatory network to be inferred. Eisen et al., (1998) applied a variant of the hierarchical average-link clustering algorithm to identify groups of co-regulated yeast genes. Chu et al., (1998) applied clustering to yeast genes to identify genes whose expression levels peak at different phases in sporulation. Golub et al.,(1999) demonstrated that clustering of the experiments can potentially be used to discover the subtypes of leukemia. Ben-Dor and Yakhini (1999) reported success with their CAST algorithm. Tamayo et al. (1999) used self-organizing maps to identify clusters in the yeast cell cycle and human hematopoietic differentiation data sets. Alon et al. (1999) split the genes through the deterministic-annealing algorithm (DAA) (Rose 1990, Rose 1998). Hierarchical clustering method is much favored by many biologists and has become the most widely-used tool in gene expression data analysis (Iyer et al.,1999, Perou et al.,1999, Alizadeh 2000). Number of clustering algorithms such as graph-based algorithms (Ben-Dor et al., 1999, Hartuv and Shamir, 2000, Shamir and Sharan, 2000, Xu et al, 2002), model-based algorithms (Fraley and Raftery 1998, Ghosh and Chinnaiyan 2002, McLachlan et al. 2002, Yeung et al 2001) and different algorithms (Shamir and Sharan 2000; Zhang et al. 2001, Smet et al., 2002; Zheng et al., 2003) are proposed for gene expression data. Different clustering algorithms are based on different clustering criteria and the performance of each clustering algorithm varies with different data sets. For example, for a gene expression data set in which the number of clusters is unknown, CAST or CLICK may be a better choice, while k-means may be the appropriate choice if the number of clusters in the data set is known in advance (Jiang et al 2004) and there is still no single clustering algorithm that finds appropriate clusters in microarray
data (Jiang et al. 2004, Ma et al. 2006). Most of the clustering algorithms are with two main problems. The first problem is that most clustering algorithms request the users to specify some parameters in advance. In real applications, however, it is hard for biologists to determine the suitable parameters manually. Thus, an automated clustering method is required. The second problem is that almost all clustering algorithms aim to produce the clustering results based on the input parameters and their own criteria. Hence, they are incapable of producing the optimal clustering result.

**Challenges in k-means algorithm**

k-means is the most popular partitional clustering technique for its efficiency and simplicity in clustering large data sets (Forgy 1965; Macqueen 1967; Hartigan and Wong 1979; Lloyd 1982). The k-means was voted as one of the top ten algorithms in data mining (Wu et al., 2008) and still is the most dominant clustering algorithm (Jain 2010). However, it also has several drawbacks as a gene-based clustering algorithm. First, the number of gene clusters in a gene expression data set is usually unknown in advance. To detect the optimal number of clusters, users usually run the algorithms repeatedly with different values of k, and compare the clustering results. For a large gene expression data set which contains thousands of genes, this extensive parameter fine-tuning process may not be practicable. Second, the k-means algorithm is sensitive to noise or outliers (Sherlock 2000). Several clustering algorithms (Ralf-Herwig et al. 1999, Heyer et al. 1999, Smet et al. 2002) are proposed to overcome the drawbacks of the k-means algorithm. These algorithms typically use some global parameters to control the quality of resulting clusters (e.g., the maximal radius of a cluster and/or the minimal distance between clusters), (Herrero et al. 2001, Tomida et al. 2002). There are two major issues in the application of partitional based (k-means type) algorithms in cluster analysis.

- Sensitive to the initial centroids or seeds
- Produce different results in different independent runs
- Number of clusters is known to be in advance

We confined here to k-means type algorithms, and made an attempt to propose algorithms to overcome the drawbacks of k-means algorithms.

**Initial Seed Selection Algorithms for k-means**

The resulting set of clusters in k-means type algorithms, both their number and
their centroids, depends on the specified choice of initial starting point values because the algorithm often converge at a local optimum (Anderberg, 1973). Therefore selecting a good set of initial seeds is very important. There is no commonly accepted or standard way to determine either the initial starting point values or the number of clusters. Many researchers introduced some methods to select good initial centers (Bradley and Fayyad, 1998). Inappropriate choice of number of clusters (Pham et al., 2004) and bad selection of initial seeds may yield poor results and may take more number of iterations to reach final solution. One of the first schemes of centroids initialization was proposed by Ball and Hall (1967). A similar approach is also provided by Tou and Gonzales (1977) under the name Simple Cluster Seeking (SCS) and is adopted in the FACTCLUS procedure. The SCS method is as follows:

1. Initialize the first cluster centroid with the first input
2. Select a point as a new seed if it is d, distance apart from the all selected seeds. Stop when k seed clusters are initialized
3. After scanning all input samples, if there are less than k seed clusters generated and then decrease d and repeat 1-2

The SCS and the method suggested by Ball and Hall are sensitive to the parameter d and the presentation order of the inputs. Astrahan (1970) suggested using two distance parameters, d1 and d2. This method first computes the density of each point in the dataset, which is given as the number of neighboring points within the distance d1 and it then sorts the data points according to decreasing value of density. The highest density point is chosen as the first seed. Subsequent seed point are chosen in order of decreasing density subject to the condition that each new seed points be at least at a distance of d2 from all other previously chosen seed points. This step is continued until no more seed points can be chosen. Finally, if more than k seeds are generated from the above approach, hierarchical clustering is used to group the seed points into the final k seeds. The drawback in this approach is that it is very sensitive to the values of d1 and d2 and requires hierarchical clustering. In the worst case it requires \(n^2 \log n\) time complexity. Kaufman and Rousseeuw (1990) introduced a method that estimates the density through pair wise distance comparison and initializes the seed clusters using the input samples from the areas with high local density. A notable drawback of the method lies in its computational complexity. Given n input samples, at least \(n(n-1)\) distance calculation are required. This could be much more time consuming than k-Means itself.
when \( n \) is large. Katsavounidis et al. (1994) suggested a parameter less approach, which is called as the KKZ method based on the initials of all the authors. KKZ chooses the first centers near the “edge” of the data, by choosing the vector with the highest norm as the first center. Then, it chooses the next center to be the point that is farthest from the nearest seed in the set chosen so far. This method is very inexpensive (\( O(kn) \)) and is easy to implement. It does not depend on the order of points and is deterministic by nature as single run suffices to obtain the seeds. However, KKZ is sensitive to outliers, since it is selecting farthest point from the selected centroids. Bradley and Fayyad (1998) proposed an initialization method that is suitable for large datasets. The main idea of their algorithm is to select \( m \) subsamples from the data set, apply the k-means on each subsample independently, keep the final \( k \) centers from each subsample provided that empty clusters are not to be allowed, so they obtained a set contains \( mk \) points, then they apply the k-means on this set \( m \) times; at the first time, the first \( k \) points are the initial centers. At the second time, the second \( k \) points are the initial centers and so on. And the algorithm returns the best \( k \) centers from this set. They use 10 subsamples from the dataset, each of size 1% of the full dataset size. Finally, a last round of k-means is performed on this dataset and the cluster centers of this round are returned as the initial seeds for the entire dataset. **This method generally performs better than k-means and converges to the local optima faster.** However, it still depends on the random choice of the subsamples and hence, may obtain a poor clustering. More recently, Arthur and Vassilvitskii (2007) proposed the k-means++ approach, which is similar to the KKZ method. Note that due to the random selection of first seed and probabilistic selection of remaining seeds, different runs have to be performed to obtain a good clustering in k-means++. Deelers and Auwatanamongkol (2007) proposed a seed selection algorithm that requires parameters in advance.

**Clustering algorithms to find number of clusters**

Despite the widespread use of clustering algorithms in gene expression data analysis (Tavazoie et al., 1999, Yeung et al. 2001; Dembele and Kastner, 2003; Sharan et al., 2003, Jiang et al. 2005) selection of clustering parameters continues to be a challenge (Tseng et al. 2005). This section contains some of the related work regarding finding clusters automatically and is reviewed here.
Fogel et al. (1996) and Sarkar et al. (1997) are proposed an approach to dynamically cluster a data set using evolutionary programming, where two fitness functions are simultaneously optimized: one gives the optimal number of clusters, whereas the other leads to a proper identification of each cluster’s centroid.

In the clustering literature (Pal and Bezdek 1995) the number of clusters in the data set is in the range from 2 to $\sqrt{m}$, where $m$ is total number of points. In many cases, the optimal specification of number of clusters, $k$, is difficult especially if there is inadequate biological understanding of the system. A suboptimal specification of number of clusters can generally result in misleading results – either all classes may not be identified or spurious classes may be generated (Bezdek and Pal 1998). While the correct number of clusters can be identified by visual inspection in some cases, in most gene expression datasets, the data dimensions are too high for effective visualization. Hence, methods that find the optimal number of clusters are essential. Several methods have been proposed for finding the number of clusters in data. The popular methods evaluate the partition using a metric and optimize it as a function of number of clusters. These methods are presented by Milligan and Cooper (1985), Halkidi et al., (2001) and Handl et al., (2005). Lee and Antonsson (2000) used an evolutionary method to dynamically cluster a data set. Some recent methods recommended for gene expression data analysis are Tibshirani et al. (2001) proposed the gap statistic that measures the difference between within-cluster dispersion and its expected value under the null hypothesis. Thus the $k$ that maximizes the difference is selected. Since the gap statistic uses within-cluster sum of squares around the cluster means to evaluate the within cluster dispersion, this method is suitable for compact, well separated clusters. Dudoit and Fridlyand (2002) proposed a prediction based re-sampling method for finding the number of clusters. For each value of $k$, the original data is randomly divided into training and testing sets. The training data is used to build a predictor for predicting the class labels of the test set. The predicted class labels are compared to that obtained by clustering of test data using a similarity metric. This value is compared to that expected under an appropriate null distribution. The $k$ for which the evidence of significance is the largest is selected. Ben-Hur et al. (2002) proposed a similar re-sampling approach where two random subsets (possibly overlapping) are selected from the data. The two random subsets are subsequently clustered independently and the similarity between the resulting partitions is measured. The similarity is measured for multiple runs and its
distribution is visualized for each \( k \). **The optimal number of clusters is selected where transition from high to low similarity occurs in the distribution.** Dudoit and Fridlyand (2002) and Ben-Hur *et al.* (2002) assume that the sample subset can represent the inherent structure in the original data which may not be true for small clusters. Furthermore, the user has to manually locate the transition in Ben-Hur *et al.* approach. Kullback Ying-Yang proposed a unified algorithm for both unsupervised and supervised learning (Guo et.al 2002), which provides a reference for solving the problem of selection of the cluster number. Recently, Bolshakova and Azuaje (2003) employed Silhouette (Rousseeuw 1987), Generalized Dunn's index, and Davies Bouldin index (1979) on gene expression data. These methods use the intra- and inter-clusters distances to identify the best partition. Discovering an optimal number of clusters in a large data set is usually a challenging task. Cheung (2005) studied a rival penalized competitive learning algorithm (Xu 1996, 1997) that has demonstrated a very good result in finding the cluster number. The algorithm is formulated by learning the parameters of a mixture model through the maximization of a weighted likelihood function. In the learning process, some initial seed centers move to the genuine positions of the cluster centers in a data set, and other redundant seed points will stay at the boundaries or outside of the clusters. Recently Swagatam Das and Ajith Abraham (2008) proposed an Automatic Clustering using Differential Evolution (ACDE) algorithm by introducing a new chromosome representation. The majority of these methods to determine the best number of clusters may not work very well in practice. The clustering algorithms are required to be run several times for good solution, and model-based methods, such as cross-validation and penalized likelihood estimation, are computationally expensive. In general, cluster validation is easier when the underlying clusters are well separated. But, most cluster validation methods lead to suboptimal results when inter- and intra-cluster distances vary largely.

**Cluster validation Methods**

Silhouette (Rousseeuw 1987), Generalized Dunn's index, and Davies Bouldin index (1979) are the various cluster validity measures to validate clustering structures of various algorithms. Despite the vast amount of knowledge available in the clustering validation of (Everitt 1993; Hansen Jaumard 1997; Hartigan 1975; Jain *et al.*1999; Mirkin 1996, Rice 1996) gene expression data provide unique challenges, in particular with respect to validation criteria. Good results can be obtained by incorporating the validation criteria
Assessing the clustering results and interpreting the clusters found are as important as generating the clusters (Jain and Dubes 1988). In much of the published clustering work on gene expression, the success of clustering algorithms is assessed by visual inspection using biological knowledge (for example, (Eisen et al., 1998). To the best of our knowledge, there is no systematic data-driven method to quantitatively evaluate gene expression clustering results. Furthermore, the clusters obtained by different clustering algorithms can be remarkably different. The evaluation criteria that help biologists to find the most valuable clusters from microarray data sets is different from the evaluation criteria in other clustering applications. The criterion for evaluating a co-expressed gene cluster needs to be able to identify whether the genes in the same cluster have the similar biological function. The Gene Ontology (GO) project (http://www.geneontology.org) evaluation tools, the Yeast genome GO term finder (http://db.yeastgenome.org/cgi-bin/GO/goTermFinder) and FatiGO evaluate the biological significance of the resulting clusters submitted by users in terms of biological processes, gene functions, and cellular components. These tools use hypergeometric distribution to compute P value to determine the statistical significance of the association of a particular GO term with a group of genes in a cluster, which evaluates whether the clusters have significant enrichment in one or more function groups (Wang et al. 2002, Liu and Wang 2003, Alsharour et al 2004, Zhao et al. 2008, Suresh et al. 2009).

1.7 Proposed Research Investigations

This dissertation made an attempt to find

- the optimal initial centroids for k-means type algorithms
- clustering algorithm insensitive to outliers
- optimal values for the parameters to be supplied in advance
- the best clustering solution
- best evaluation method to study the performance of clustering algorithm
- a clustering algorithm which determines optimal number and optimal clusters automatically
• optimal, robust, single clustering solution that scales for large data sets
• number of gene groups existed in the data set
• the co-expressed gene groups in the data set
• the biological enrichment of a group of genes clustered together

in cluster analysis of real data sets including microarray data sets.

1.8 Limitations of Study

The present work proposes certain new clustering algorithms to identify optimal number of clusters, their centroids, and formulate appropriate clusters automatically from the given gene expression data, as an extension to k-means. The algorithms proposed are find hard clustering solutions. The algorithms incorporate a cluster validity measure in identifying optimal clusters automatically, therefore the results of the algorithms may depend on the selected validity measure and the clustering structure also depends on the validity criteria. The genes grouped into a cluster as a resultant of implementation of optimal algorithm is to be validated using domain biological knowledge before releasing them for automation to a software company.

1.9 Thesis Overview

The results of the above mentioned investigations are documented into six chapters. In the first chapter a general introduction to microarray clustering, focus of study, challenges in clustering process, statement of the problem, survey of literature, proposed research investigations and limitations of study are presented.

The second chapter is devoted to present an overview of the applications of clustering, and the statistical methods in bioinformatics are included in the third chapter.

The existing partitional clustering algorithms, and their pitfalls, newly proposed algorithms, various synthetic, real multivariate and public microarray data sets will form the content of chapter four.

Penultimate chapter contains the comparative study of the experimental results of the proposed algorithms, performance measures and efficiency when compared to
existing algorithms. The results are tabulated and comments on the results presented in the tables are included.

Conclusions and further scope of the work is presented in the last chapter.