Trend Analysis of Telecom data using Unsupervised Neural Networks

Developed by:

Muhammad Asif (210-FAS/MSCS/F04)
Atif Hameed Bhatti (193-FAS/MSCS/F04)

Supervised By:

Mr. Shakeel Ahmad

Faculty of Basic & Applied Sciences
Department of Computer Science
International Islamic University, Islamabad
(2008)

بسم الله الرحمن الرحيم

In the name of Almighty Allah,
The most Beneficent, the most Merciful.

Department of Computer Science,
International Islamic University, Islamabad.

June 30 , 2008

Final Approval

It is certified that we have read the thesis, titled "*Trend Analysis of Telecom Data using Unsupervised Neural Networks*" submitted by **Muhammad Asif,** Reg. No. 210-FAS/MSCS/2004 and **Atif Hameed Bhatti**, Reg. No. 193-FAS/MSCS/2004. It is our judgment that this thesis is of sufficient standard to warrant its acceptance by the International Islamic University, Islamabad, for the Degree of Master of Science.

Committee

External Examiner

Dr. Sajjad Mohsin

Associate Professor
Department of Computer Science, COMSATS Institute of Information Technology, Islamabad.

Internal Examiner

Syed Muhammad Saqlain

Assistant Professor
Department of Computer Science, International Islamic University, Islamabad.

Supervisor

Mr. Shakeel Ahmad

Assistant Professor
Department of Computer Science, International Islamic University, Islamabad.

Dedication

"What we cannot speak about we must pass over in silence."
L. Wittgenstein

To our parents

A dissertation submitted to the

Department of Computer Science,

International Islamic University, Islamabad

In partial fulfillment of the requirements

For the award of the degree of

Master of Science

Declaration

We, the undersigned hereby declare that this research, *Trend Analysis of Telecom data using Unsupervised Neural Networks*, is our own work. It is further declared that this work, neither in part nor in full, has been copied from any source of information. Also, it has not been submitted for any degree or examination in any other university, and that all the sources we have used or quoted have been indicated and acknowledged by complete references.

Muhammad Asif
Reg. No. 210-FAS/CS/MS/2004

Atif Hameed Bhatti
Reg. No. 193-FAS/CS/MS/2004

Acknowledgment

All praise to the Almighty Allah (Tabrak Wa'tala), the most Merciful and the most Gracious one --- without Whose help and blessings, we would not have been able to complete this work.

Working on what is now presented in this report has been a great experience for us. This is largely due to the truly outstanding support that we have received from our supervisor ***Mr. Shakeel Ahmad***. We are most grateful for the long time he has spent discussing the project with us and the advice and guidance he has given us. His encouragement and enthusiasm have been invaluable.

Further, we would like to express our profound gratitude goes to our respected teacher, ***Dr. Syed Afaq Hussain***, for his valuable contribution and expertise. His constant motivation, unflagging efforts and uninvolved words of wisdom ever proved a lighthouse for us. He was an excellent reference point for us. His patience, motivation and guidance contributed significantly to the completion of this dissertation.

Acknowledgement is also due to our respected friend, ***Mansoor Hasan Khan***, for his valuable help, support and encouragement throughout our project.

Project in Brief

Project Title	:	Trend Analysis of Telecom Data using Unsupervised Neural Networks
Objective	:	To investigate the unsupervised learning potential of neural networks for the trend analysis of calls made by users over a period of time in a mobile telecommunication network.
Undertaken By	:	Muhammad Asif, Atif Hameed Bhatti
Supervised By	:	Mr. Shakeel Ahmad
Technologies Used	:	MATLAB 7.0
System Used	:	Pentium® IV, 1GB RAM
Operating System Used	:	Windows XP SP2
Date Started	:	August, 2007
Date Completed	:	March, 2008

Abstract

The primary aim of this thesis is to investigate the unsupervised learning potentials of novel neural networks for the trend analysis of calls made by users over a period of time in a mobile telecommunication network. Specifically, this study provides a comparative analysis and application of Self Organizing Maps (SOM), Learning Vector Quantization (LVQ) and Adaptive Resonance Theory (ART) competitive neural networks algorithms to user call data records in order to conduct a descriptive data mining on users call patterns.

Huge amounts of data are being collected as a result of the increased use of mobile telecommunications. Insight into information and knowledge derived from these databases can give operators a competitive edge in terms of customer care and retention, marketing and fraud detection.

Unsupervised neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyze. Unsupervised learning is very useful and remarkable technique when dealing with large volumes of raw data with little or no knowledge of the inter relation between the various fields in a vector. This network can create its own organization or representation of the information it receives during learning time.

Table of Contents

Chapter No. & Title			Page #
1	**Introduction**	....	1
	1.1 Overview		1
	1.2 Literature Survey		2
	1.3 Problem Formulation		7
	1.4 Motivation		8
	1.5 Objective of the Project		9
	1.6 Contribution		9
2	**Overview of the Techniques**	....	11
	2.1 Trend Analysis		11
	2.1.1 Importance of Trend Analysis		11
	2.1.2 Benefits of Trend Analysis		11
	2.2 Data Mining		12
	2.2.1 Data Mining Process		12
	2.2.2 Data Mining Approaches		14
	2.3 Clustering		14
	2.3.1 Clustering Methods		15
	2.4 Artificial Neural Networks		15
	2.4.1 Supervised Learning		17
	2.4.2 Unsupervised Learning		18
	2.5 The Self-Organizing Map (SOM)		18
	2.6 The Adaptive Resonance Theory (ART)		19
	2.7 The Learning Vector Quantization (LVQ)		20

3	**Requirement Analysis**	....	**22**
3.1	Introduction		22
3.2	Problem Statement		22
3.3	Problem Scenarios		23
4	**System Design**	....	**25**
4.1	Introduction		25
4.2	Reference Architecture		25
4.3	Design Methodology		28
4.4	Algorithms		28
4.4.1	Self Organizing Map (SOM)		29
4.4.2	The Adaptive Resonance Theory (ART)		31
4.4.3	The Learning Vector Quantization (LVQ)		36
5	**Implementation and Results**	....	**38**
5.1	Data Selection and Preparation		38
5.2	Model Building		39
5.2.1	Self Organizing Map (SOM)		39
5.2.2	Learning Vector Quantization (LVQ)		40
5.2.3	Adaptive Resonance Theory (ART)		40
5.3	Evaluation		40
5.3.1	Clustering Evaluation		40
5.3.1.1	Clustering of Revenue Data		41
5.3.1.2	Clustering of Time of call Data		47
5.3.2	Trend Analysis		53
5.3.2.1	Trend Analysis of Revenue Data		53
5.3.2.2	Trend Analysis of Time of Call Data		59

6 Conclusion and Outlook **64**

6.1 Conclusion 64

6.2 Future System Improvements 66

7 References **67**

List of Figures

Figure #	Name	Page #
Figure 1:	Data Mining Process	13
Figure 2:	Clustering provides an overview of the data	14
Figure 3:	A Neural Network: An interconnected group of nodes	16
Figure 4:	Artificial Neural Network classification model	17
Figure 5:	Each map node is associated with a vector of weights	19
Figure 6:	CRISP-DM (Cross Industry Standard Process for Data Mining)	26
Figure 7:	Data Mining Design Methodology	28
Figure 8a:	Typical SOM architecture	29
Figure 8b:	Typical ART2 architecture	31
Figure 8c:	Typical LVQ architecture	36
Figure 9:	SOM based Clusters for Revenue	41
Figure 10:	LVQ based Clusters for Revenue	42
Figure 11:	ART based Clusters for Revenue	43
Figure 12:	SOM based Cluster Ranges for Revenue	44
Figure 13:	LVQ based Cluster Ranges for Revenue	45
Figure 14:	ART based Cluster Ranges for Revenue	46
Figure 15:	SOM based Clusters for Time of Call	47
Figure 16:	LVQ based Clusters for Time of Call	48
Figure 17:	ART based Clusters for Time of Call	49
Figure 18:	SOM based Cluster Ranges for Time of Call	50
Figure 19:	LVQ based Cluster Ranges for Time of Call	51
Figure 20:	ART based Cluster Ranges for Time of Call	52
Figure 21:	SOM based Trends of Most Revenue generating Packages	53
Figure 22:	LVQ based Trends of Most Revenue generating Packages	54
Figure 23:	ART based Trends of Most Revenue generating Packages	55
Figure 24:	SOM based Trends of Most Revenue generating Cities	56
Figure 25:	LVQ based Trends of Most Revenue generating Cities	57
Figure 26:	ART based Trends of Most Revenue generating Cities	58
Figure 27:	SOM based Trends of Most Busy Hours	59
Figure 28:	LVQ based Trends of Most Busy Hours	60
Figure 29:	ART based Trends of Most Busy Hours	60
Figure 30:	SOM based Trends of Cities with Most Busy Hours	61
Figure 31:	LVQ based Trends of Cities with Most Busy Hours	62
Figure 32:	ART based Trends of Cities with Most Busy Hours	63

Chapter 1

INTRODUCTION

1. Introduction

1.1 Overview

Over the past decade, academic as well as commercial databases have been growing at exceptional rates. Gaining new knowledge from such databases is difficult, costly and time-consuming if done manually. It may even be impossible when the data exceeds certain limits of size and complexity. As a result, the automated analysis and visualization of massive multi-dimensional datasets has been the focus of much scientific research during the last years. The principal objective is to find regularities and relationships in the data, thereby gaining access to hidden and potentially useful knowledge. Artificial Neural Networks are a promising part of this broad field. Inspired by advances in biomedical research, they form a class of algorithms that aim to simulate the neural structures of the brain [Jost (2003)].

There are several criteria to classify Neural Networks; the two that are most important are the training mode and the feed back mechanism. The training mode could be supervised and unsupervised. In Supervised learning mode Neural Networks rely on a previously compiled and sanitized data "training set". The network then tweaks its internal "weights" such that it will try to accurately classify the largest majority of training vectors. The benefit of this technique is that it has a more generalized coverage of its problem domain. In Unsupervised learning mode, instead of being told what they should be looking for or what to report, network try to find patterns within a data set and seeks to group them according to the most relevant features. This technique is very useful when dealing with large volumes of raw data with little or no knowledge of the inter relation between the various fields in a vector [Hussam (2002)].

The Self-Organizing Map (SOM) *is a fairly well-known neural network and indeed* one of the most popular unsupervised learning algorithms. Since its invention by Finnish Professor Teuvo Kohonen in the early 1980s [Kohonen (1981); Kohonen (1983)], more than 4000 research articles have been published on the algorithm, its visualization and applications. The maps comprehensively visualize natural groupings and relationships in the data and have been successfully applied in a broad spectrum of research areas ranging from speech recognition to financial analysis. The Self-Organizing Map performs a non-linear projection of multidimensional data onto a two dimensional display. The mapping is topology-preserving, meaning that the more alike two data samples are in the input space, the closer they will appear together on the final map. This allows the user to identify "clusters", i.e. large groupings of a certain type of input pattern. Further examination may then reveal what features the members of a cluster have in common [Jost (2003)].

Adaptive Resonance Theory (ART) architecture is another unsupervised neural network that carries out stable self-organization of recognition codes for arbitrary sequence of input pattern. . It allows increase in number of clusters only if required. Adaptive Resonance Theory first emerged from an analysis of the instabilities inherent in feed forward adaptive coding structure [Grossberg (1976a), 1976b)]. More recent work has led to the development of three classes of ART neural network architecture, specified as system differential equations: ART1 and ART2 [Carpenter & Grossberg (1987a), (1987b)]. By especially ART2 self-organizes recognition categories for arbitrary sequences of either binary or analog inputs. ART2 is designed to perform for continuous-valued input vectors the same type of tasks as ART1 does for binary input vectors. The differences between ART2 and ART1 reflect the modifications need to accommodate patterns with continuous-valued components. The more complex F1 field of ART2 is necessary because continuous-valued input vector may be arbitrarily close together. The F1 field in ART2 includes a combination of normalization and noise suppression, in addition to the comparison of the bottom-up and top-down signals needed for the reset mechanism. [Carpenter et al. (1991)].

1.2 Literature Survey

Literature Survey is an important and unavoidable part of research, as without literature survey we cannot understand the scenario that till what point the researchers have reached and what are the loopholes in the topic and what can be enhanced in that area.

1.2.1 Mark W. Craven, Jude W. Shavlik, 1998: Using Neural Networks for Data Mining [Craven and Shavlik (1998)]

In this paper, authors describe neural network learning algorithms that are able to produce comprehensible models and that do not require excessive training times.

Contribution

The authors have presented algorithms that are able to extract symbolic rules from trained neural networks, and algorithms that are able to directly learn comprehensible models. The authors have used competitive learning algorithm and argued that it is the most used algorithm for data mining. It is an on-line algorithm, meaning that during training it updates the network's weights after every example is presented. Thus online makes it more suitable for very large data sets as compared to statistical model K-means clustering.

1.2.2 Portia A. Cerny, 2002: Data mining and Neural Networks from a Commercial Perspective [Cerny (2001)]

In this paper, the authors have discussed that Neural networks have proven useful for a wide range of Commercial Applications including fraud detection, telecommunications, medicine, marketing, bankruptcy prediction, insurance etc. They have discussed the following advantages:

High Accuracy, Noise Tolerance, Independence from prior assumptions, Ease of maintenance, Neural networks can be implemented in parallel hardware, Neural networks performance can be highly automated, minimizing human involvement.

1.2.3 Jost Schatzmann, 2003: Using Self-Organizing Maps to Visualize Clusters and Trends in Multidimensional Datasets [Jost (2003)]

The author tells about the use of SOM algorithm for clustering and information visualization in the past. He argues that although work has been done in this field but a system has been lacking that combines the fast execution of the algorithm with powerful visualisation of the maps and effective tools for their interactive analysis.

Contribution

The author has given system which is Java implementation of SOM algorithm which is controlled through a graphical user interface. It unites many successful ideas that were only implemented independently so far. In addition, it offers new features that have not been published elsewhere such as the option to isolate selected clusters.

1.2.4 Pavel Berkhin, 2004: Survey of Clustering Data Mining Techniques [Pavel (2002)]

The author has carried out a survey that focuses on clustering in data mining. The authors found out that unlike traditional hierarchical methods, in which clusters are not revisited after being constructed, relocation algorithms gradually improve clusters. With appropriate data, this results in high quality clusters. On the other hand, K-means method provides better results in case of numerical values. Centroids have the advantage of clear geometric and statistical meaning.

1.2.5 Olusola Adeniyi Abidogun, 2005: Call Pattern Analysis with Unsupervised Neural Networks [Abidogun (2005)]

In this paper the authors have noted that unsupervised neural networks can mainly be used in exploratory data analysis. They act on unlabelled data in order to extract an efficient internal representation of the structure implicit in the data distribution. The author's main focus has been to investigate the unsupervised learning potentials of two neural networks - Self-Organizing Maps (SOM) and Long Short-Term Memory (LSTM) recurrent neural networks for the profiling of calls made by users over a period of time in a real mobile telecommunication network in order to conduct a descriptive data mining on the users call patterns.

Contribution:
The authors showed that LSTM recurrent neural network algorithm provides a better discrimination of call patterns than the SOM algorithm in terms of long time series modeling and group them according to certain features. The ordered features were later interpreted and labeled according to specific requirements of the mobile service provider. Thus, suspicious call behaviors' were isolated within the mobile telecommunication network.

1.2.6 Gary M. Weiss, 2005: Data Mining in Telecommunications [Gary (2005)]

In this paper the author has described how data mining can be used to uncover useful information buried within telecom data sets. There are three main sources of telecommunication data (call detail, network and customer data). Most common data mining applications for telecommunication data are 1) Fraud Detection, 2) Customer profiling and 3) Network Fault Detection. One Big problem with Call Detail Data is that it is impossible to use all of it for data mining, due to immense storage and processing challenges. So important fields need to be identified and used to find often rare patterns in the data.

1.2.7 Martin Hynar, Michal Burda, and Jana Sarmanov, 2005: Unsupervised Clustering with growing self-organizing neural network – a comparison with non-neural approach [Martin et al. (2005)]

In this paper the authors have stated that non neural approaches like K-means methods are too static in a particular point of view; at the beginning we have to specify the number of expected clusters and the algorithm is then responsible to find them in the data set. But, what if we could not determine this number? Therefore they have preferred SOM as SOM has its topology preserving behaviour. This means, that SOM tries to adapt weights of neurons to cover the most dense regions and therefore SOM naturally finds data clusters. The limitation of SOM lies in fact that it is designed to have number of neurons specified as the input parameter and immutable during the learning process. The most important feature of such neural networks is their natural ability to find dense areas in the input space.

1.2.8 Gail A. Carpenter, Stephen Grossberg, David B. Rosen, 1991: ART2-A: An Adaptive Resonance Algorithm for Rapid Category Learning and Recognition [Carpenter et al. (1991)]

The authors have discussed the importance and its placement for large problem solving. They have presented the following advantages of using ART2.

Advantages:
1) ART 2 is better suited for large problem solving on conventional computers. It can also be efficiently implemented on parallel systems for even greater performance boost.

2) It retains both distant and recent past information. It can learn fast with both Linear and Non-linear STM (Short term Memory) feedback

3) At Intermediate level, it is good for LTM

1.2.9 Sauli Kivikunnas, 1998: Overview of Process Trend Analysis Methods and Applications [Kivikunnas (1998)]

This paper surveys trend analysis methods and their applications in process industry. Basic principles of several methods are presented together with known applications in process monitoring, diagnosis and control. In the first place, the methods presented analyse process measurements or some calculated quantities as time series using various pattern recognition methods. Also computationally much lighter methods such as linear regression based methods are included. Researchers with different background, for example from pattern recognition, digital signal processing and data mining, have contributed to the trend analysis development and have put emphasis on different aspects of the field

1.2.10 Hilary K. Browne,William A. Arbaugh,John McHugh,William L. Fithen, 2001: A Trend Analysis of Exploitations [Hillary et al. (2001)]

The authors have discussed different types of security issues in computer systems and how much they are vulnerable to such exploitations.
CERT/CC data is the best available source for an analysis of this type.

Contribution:
In this paper, authors have presented a statistical model that relates the rate at which intrusions accumulate, and provided evidence to support it. The result is a model that assists in predicting the severity of an exploitation cycle. The existence of a severity predictor allows incident handling organizations to plan and staff accordingly. Additionally, the knowledge of the severity of an incident can assist operational organizations in performing more effective risk management. The model, indicates that each of the vulnerabilities that we have studied accumulate in a similar, and near linear, fashion. Identifying and validating the model requires a regression analysis on the intrusion data for each vulnerability.

Given the results of the regression analyses above, a linear regression model using a square root transformation on time appears to provide very good predictive power for the accumulation of security vulnerability incidents following the release of a script for the vulnerability.

Advantage:
Such analysis permits the organization to plan it's staffing requirements rather than reacting. Operational organizations can benefit from the knowledge of the severity of

continuing incidents. For instance, most operational organizations test vendor supplied patches prior to deployment to ensure that the fix for the vulnerability does not produce unwanted side effects. In the case of security related patches, a time-bar is usually established as to when the patch must be deployed.

1.2.11 K. Rajaraman and Ah-Hwee Tan, 2001: Topic Detection, Tracking and Trend Analysis Using Self-organizing Neural Networks. [Kanagasabi and Ah-Hwee (2001)[

The authors have used ART for the purpose of detection of text streams and then their trend analysis is carried out to predict for future. ART formulates recognition categories of input patterns by encoding each input pattern into a category node in an unsupervised manner. Thus each category node encodes a cluster of patterns. In other words, each node represents a topic.

Trend analysis

The topic detection and tracking setup together with the time ordering of the documents provides a natural way for topic-wise focused trend analysis. In particular, for every topic, suppose we plot the number of documents per segment versus time. This plot can be thought of as a trace of the evolution of a topic. The `ups' and `downs' in the graph can be used to deduce the trends for this topic.

1.2.12 R. Inokuchi and S. Miyamoto, 2004: LVQ Clustering and SOM Using a Kernel Function. [Inokuchi and Miyamoto (2004)]

The authors have discussed clustering algorithm based on Learning Vector Quantization (LVQ) using a kernel function (K-LVQC). They argue that when standard methods are used for clustering, these give linear cluster boundaries. On the other hand, real world situations request us clusters having nonlinear boundaries. To obtain nonlinear classification boundaries, a powerful method is the use of kernel functions, where original data are mapped into a higher dimensional feature space; data are classified linearly in the feature space but in the original data space the boundary appear nonlinear. Moreover an SOM using the kernel visualizes how the data is separated in the feature space.

Four other methods of fuzzy cmeans, i.e., the standard fuzzy c-means sFCM, entropy based fuzzy c-means eTCM without a kernel function, K-sFCM (the standard fuzzy c-means with the kernel), and K-eFCM (the entropy-based fuzzy e-means with the kernel) were tested. Results showed that the number of misclassified objects found by K—LVQC were as good as other methods and the use of the kernel is effective in the real data sets. Also processor time taken showed that K-eFCM is time consuming while K-LVQC is much more efficient, while the classification capability is not worse than K-eFCM and eFCM.

1.2.13 Dolnicar, S., & Leisch, F., 2001: Behavioral market segmentation of binary guest survey data with bagged clustering. [Dolnicar and Leisch. (2001)]

The authors have discussed the problems of traditional hierarchial and partitioning approaches and therefore presented Bagged clustering technique to overcome the problems. They have applied bagged cluster algorithm to a binary data set from tourism marketing.

Drawbacks in previous techniques

Hierarchical clustering techniques require the data sets to be rather small, as all pairwise distances need to be computed in every single step of the analysis. Partitioning approaches like learning vector quantization (LVQ) typically give less insight into the structure of the data, as the number of clusters has to be specified a-priori and solutions for different number of clusters can often not be easily compared.

Contribution

Bagged clustering overcomes these difficulties by combining hierarchical and partitioning methods. Clusters can be split into sub-segments, each branch of the tree can be explored and the corresponding market segment identified and described. Learning vector quantization was used as base method. For stability checks, standard K-means and LVQ with bagged versions were compared. K-Means and LVQ were independently repeated 100 times using K = 3 to 10 clusters.

Results

Results showed that Bagging considerably increases the mean agreement of the partitions for all number of clusters while simultaneously having a smaller variance. Hence, the procedure stabilizes the base method. Results also showed that LVQ is more stable than K-Means on the given binary data set.

1.3 Problem Formulation

During Literature Survey, we have seen that although SOM has been used for Trend Analysis of different types of data including Telecom but LVQ and ART have not been used as a basis where unsupervised learning in a dynamic model fashion is applied to Telecom data for trend analysis. Also comparative study of these unsupervised neural network algorithms has not yet been carried out. Hence the effectiveness and comparative analysis of these different techniques has not yet been explored in detail. Therefore, until and unless we implement these techniques on the data, we will be unable to predict accurately the selection and application of individual algorithms in different cases on telecom data.

1.4 Motivation

Before describing importance of unsupervised neural network methods for data mining, we first make an argument for why one might want to consider using neural network for the task. The essence of the argument is that, for some problems, neural networks provide a more suitable *inductive bias* than competing algorithms. Let us briefly discuss the meaning of the term inductive bias. Given a fixed set of training examples, there are infinitely many models that could account for the data, and every learning algorithm has an inductive bias that determines the models that it is likely to return. There are two aspects of the inductive bias of an algorithm: its *restricted hypothesis space bias* and its *preference bias*. The restricted hypothesis space bias refers to the constraints that a learning algorithm places on the hypothesis that it is able to construct. For example, the hypothesis space of a perceptron is limited to linear discriminant functions. The preference bias of a learning algorithm refers to the preference ordering it places on the models that are within its hypothesis space. For example, most learning algorithms initially try to fit a simple hypothesis to a given training set and then explore progressively more complex hypothesis until they find an acceptable fit.

One appealing aspect of many neural network learning methods, however, is that they are *on-line* algorithms, meaning that they update their hypothesis after every example is presented. Because they update their parameters frequently, on-line neural network learning algorithms often converge much faster than *batch* algorithms. This is especially the case for large data sets. Often, a reasonably good solution can be found in only one pass through a large training set. For this reason, we argue that training-time performance of neural network learning methods may often be acceptable for data mining tasks, especially given the availability of high performance, desktop computers. [Craven and Shavlik (1998)]

Unsupervised neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques [Dimitrios and Christos (1997)]. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyze. Unsupervised learning is very useful and remarkable technique when dealing with large volumes of raw data with little or no knowledge of the inter relation between the various fields in a vector. This network can create its own organization or representation of the information it receives during learning time.

As we are doing Trend analysis of telecom data, where the data is huge and structure is unknown, therefore unsupervised learning helps in extracting different patterns and come out with trends in the form of cluster building which are helpful for the telecommunication industries to make and adjust their future policies on the basis of the results. Hence this is the basis of motivation for the selection of unsupervised neural network algorithms for finding trends in the Telecom data.

1.5 Objective of the Project

The objective is to build a generic system based on the unsupervised neural network algorithms for the trend analysis of Telecom data and carry out a comparison between them giving their efficiencies. The system must be capable of generating and visualizing graphs, thus displaying the trend and cluster relationships hidden in the data.

1.6 Contribution

The contribution of our work is two-fold. Firstly we have designed a system in MATLAB by implementing SOM, LVQ and ART algorithms. Implementation has been done using two scenarios. In the first scenario, the clusters are generated on the basis of user revenues while in the second scenario, network usage is considered. Each algorithm takes input and then categorizes it into different clusters based on the above three NN techniques. These clusters form the basis for data and trend analysis. The graphs generated portray the trends hidden in the data within a cluster as well as across the clusters.

Secondly, we have provided a comparative analysis of the outputs/results generated by Self-Organizing Maps (SOM), Learning Vector Quantization (LVQ) and Adaptive Resonance Theory (ART) unsupervised neural networks algorithms. Our research found important characteristics and advantages of one algorithm over the other. SOM algorithm categorizes data basically with respect to first component of input vector (First component is revenue while second is call time in our case). It means that the algorithm considers only the revenue component to carry out the clustering. LVQ algorithm is fast and the degree of correctness of its categorization depends on the number and type of learning data. Also it is fast as it takes the output of SOM as its input. As far as ART algorithm is concerned, it depends on all the components of input vector (in our case two only) for the creation of clusters. It does not categorize data if input vector has only one component. We see that the choice of the parameter θ has a significant effect on the performance of the network. Also, the value of vigilance parameter ρ determines how many clusters will be formed.

Organization of the Thesis

This thesis is organized into 6 chapters.

Chapter 1 introduces the overall idea of the project, covers the Literature survey relevant to the project. It explains a comprehensive review of the published and unpublished work from secondary sources data in the areas of Data Mining, Artificial Neural Network and Telecom related research work. It sketches out the motivation that led to its initial proposal and outlines the contribution of the work presented in this report.

Chapter 2 covers the overview of the techniques relevant to the project. Starting with the more general concepts of Trend Analysis, Data Mining, Artificial Neural Networks and Clustering and then moves on to explain Self-Organizing Maps, Learning Vector Quantization and Adaptive Resonance Theory.

Chapter 3 covers the Requirement Analysis. It means that the chapter aims to specify what exactly the system is expected to deliver.

Chapter 4 covers the design of the system. Again, we start with more general design priorities and design choices before we move on to specific details of the design and implementation. In particular, the chapter covers the clustering and trend analysis with graphical representations.

Chapter 5 describes the implementation and results of some case studies. The studies demonstrate various aspects of the system in application to real-world problems.

Chapter 6 summarizes the report findings and draws a conclusion.

Chapter 2

OVERVIEW OF THE TECHNIQUES

2. Overview of the Techniques

In this chapter, we have explored different techniques upon which our thesis is based.

2.1 Trend Analysis

The term *"trend analysis"* refers to the concept of collecting information and attempting to spot a pattern, or *trend*, in the information [Trend analysis (2007)]. Although trend analysis is often used to predict future events, it could be used to estimate uncertain events in the past. Trend analysis is based on the idea that what has happened in the past gives companies an idea of what will happen in the future.

2.1.1 Importance of Trend Analysis

Timely identification of emerging trends is a key factor of success for any business. Data analysis including Trend Analysis is essential for a firm's competitive intelligence program. The ability to accurately gauge customer response to changes in business and other environmental parameters is a powerful competitive advantage.

Trend Analysis Methods are essential to running an organization's value chains and in acquiring and consolidating corporate success. It allows business users to make analytical decisions about what direction the business should target its resources on and to focus on those business processes that maximize revenue from core customers.

With the information explosion, an incredible amount of information is available to organizations. However, raw data by itself does not provide much information. It is the conversion of this raw data into significant facts, relationships, trends and patterns that could otherwise go unobserved.

2.1.2 Benefits of Trend Analysis

Trend Analysis provides an insight into the following [Trend analysis (2008)]:

- Changes and trends in customer needs and behavior, and shifts in the customers' perception of value
- Trend in price change and cost drivers for the industry and/or specific segments
- Change and evolution of the industry in terms of new entrants, and competition, threat of substitutes and relationship with buyers and suppliers
- Upcoming business models and changing best practices of the industry and related emerging sectors

- In depth analysis of long term industry, domestic and global economic cycles and trends.
- Insight into services and product purchasing trend patterns
- Analyzes common characteristics of a consumer base
- Identifying those consumers who are most likely to discontinue that service or product
- Predetermining those transactions that are most likely to be fraudulent, taking into account previous trend analyses
- Predicting in advance the products or services a person is most likely to use based on past and present trends

2.2 Data Mining

Data mining is frequently described as *"the process of extracting valid, authentic, and actionable information from large databases."*
Data Mining, or Knowledge Discovery in Databases (KDD) as it is also known, is the nontrivial extraction of implicit, previously unknown, and potentially useful information from data [Frawley et al. (1991)].

Data mining is the search for relationships and global patterns that exist in large databases but are `hidden' among the vast amount of data. These relationships represent valuable knowledge about the database and the objects in the database and, if the database is a faithful mirror, of the real world registered by the database [Holsheimer and Siebes (1994)].

Data mining derives patterns and trends that exist in data. These patterns and trends are stored and defined as a mining model. Mining models can be applied to specific business scenarios. You can use this model to extrapolate trends.

The research field of data mining has developed sophisticated methods for identifying patterns in data in order to provide insights to users. Identifying *temporal* relationships (e.g., trends) in data constitutes an important problem that is relevant in many business and academic settings, and the data mining literature has provided analytical techniques for some specialized types of temporal data [Clementine (2008)].

2.2.1 Data Mining Process

The process of data mining consists of three stages [StatSoft (2007)]:

(1) The initial exploration,
(2) Model building or pattern identification with validation/verification, and
(3) Deployment (i.e., the application of the model to new data in order to generate predictions).

Stage 1: Exploration.

This stage usually starts with data preparation which may involve cleaning data, data transformations, selecting subsets of records and - in case of data sets with large numbers of variables ("fields") - performing some preliminary feature selection operations to bring the number of variables to a manageable range (depending on the statistical methods which are being considered). Then, depending on the nature of the analytic problem, this first stage of the process of data mining may involve anywhere between a simple choice of straightforward predictors for a regression model, to elaborate exploratory analyses using a wide variety of graphical and statistical methods (see *Exploratory Data Analysis (EDA)*) in order to identify the most relevant variables and determine the complexity and/or the general nature of models that can be taken into account in the next stage.

Stage 2: Model building and validation.

This stage involves considering various models and choosing the best one based on their predictive performance (i.e., explaining the variability in question and producing stable results across samples). This may sound like a simple operation, but in fact, it sometimes involves a very elaborate process. There are a variety of techniques developed to achieve that goal - many of which are based on so-called "competitive evaluation of models," that is, applying different models to the same data set and then comparing their performance to choose the best. These techniques - which are often considered the core of predictive data mining - include: Bagging (Voting, Averaging), Boosting, Stacking (Stacked Generalizations), and Meta-Learning.

Stage 3: Deployment. That final stage involves using the model selected as best in the previous stage and applying it to new data in order to generate predictions or estimates of the expected outcome.

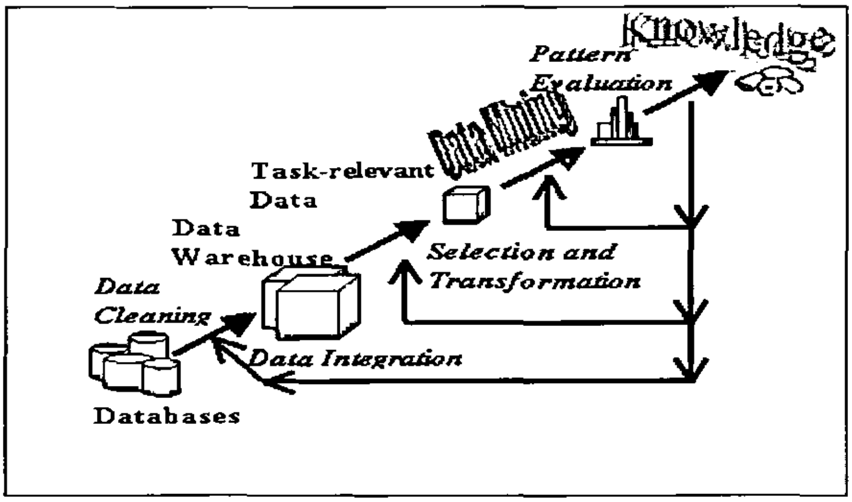

Figure 1: Data Mining Process [Frawley et al. (1991)]

2.2.2 Data Mining Approaches

Traditionally, the data mining community has focused on algorithms, which are specific techniques to approach a task. Many algorithms have been developed and extensively tested both in practice and in the academic community. Algorithms that are in wide use fall into the following broad categories:

- Classification
- Estimation
- Prediction
- Association
- Clustering

2.3 Clustering

In many practical situations, vast sets of unknown multi-dimensional data are present. Clustering is an approach to identify natural groupings of similar entries in such sets of unclassified data - often without any a priori knowledge as to what that similarity may involve. The principal idea is to partition the dataset into meaningful sub-classes, called clusters. Especially if good visualization support is available (see Figure 2), clustering can provide a helpful .first impression. of the way the data is distributed. It is therefore often undertaken as an exploratory exercise before doing further data mining.

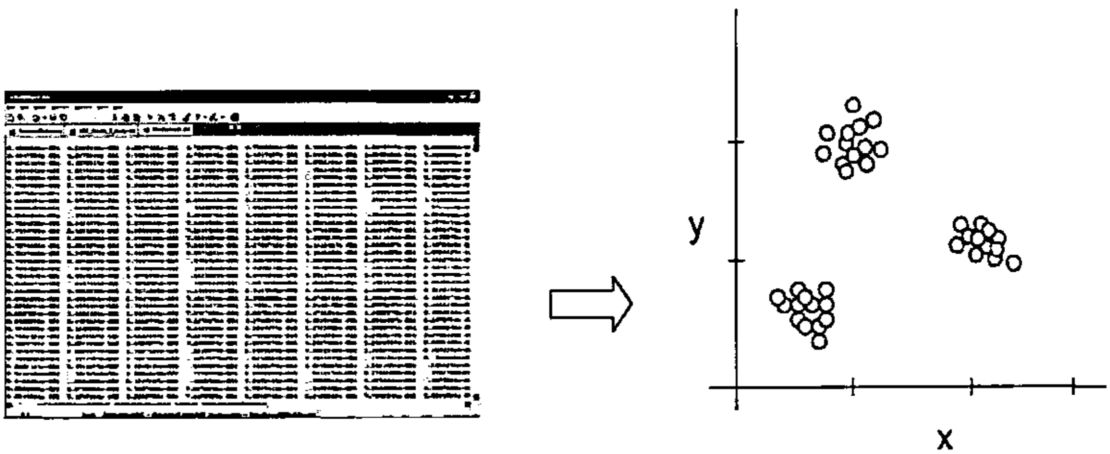

Figure 2: Clustering provides an overview of the data

2.3.1 Clustering Methods

The two major clustering methods are Partitioning and Hierarchical Clustering [Jost (2003)].

The basic concept of partitioning methods is to divide a large dataset of *n* objects into a small set of *k* clusters. Generally, some partitioning criterion is defined and given a *k*, we aim to find a partition of *k* clusters which optimizes this criterion. The criterion is often formulated as an objective function (cost function.) which is to be minimized as much as possible. K-Means, which is perhaps the most commonly used partitioning method.

In hierarchical clustering, one can either follow a divisive (top-down.) or agglomerative "bottom-up" approach. In the latter, we start with each data item in its own cluster and iteratively combine the clusters to form larger and larger ones. More precisely, we determine the two most similar clusters in each step and merge these into one. The procedure terminates when there is only a singe cluster left. Divisive methods on the other hand start with a single cluster containing all data items. During each step, the least coherent cluster is determined and split into two. This can be repeated until the number of clusters equals the numbers of items in the dataset.

2.4 Artificial Neural Networks

An artificial neural network (ANN) or commonly known as neural network (NN) is one of the methods used in Data Mining. ANN is an interconnected group of artificial neurons that uses a mathematical model or computational model for information processing based on a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network [Olubunmi (2004)].

The commonest type of artificial neural network consists of three groups, or layers, of units: a layer of "input" units is connected to a layer of "hidden" units, which is connected to a layer of "output" units. The activity of the input units represents the raw information that is fed into the network. The activity of each hidden unit is determined by the activities of the input units and the weights on the connections between the input and the hidden units. The behaviour of the output units depends on the activity of the hidden units and the weights between the hidden and output units. This simple type of network is interesting because the hidden units are free to construct their own representations of the input. The weights between the input and hidden units determine when each hidden unit is active, and so by modifying these weights, a hidden unit can choose what it represents. [Siganos and Stergiou (1997)].

In more practical terms neural networks are non-linear statistical data modeling tools. They can be used to model complex relationships between inputs and outputs or to find patterns in data.

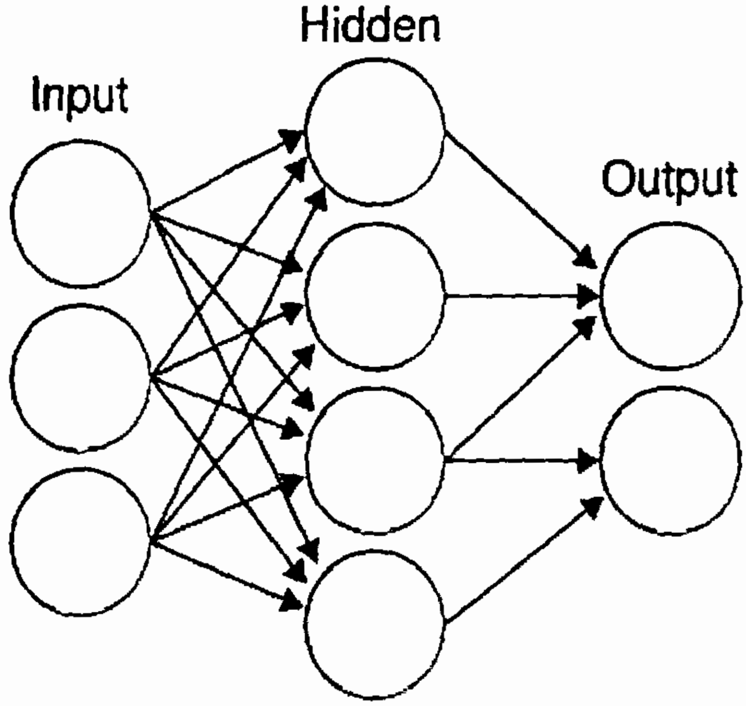

Figure 3: A Neural Network: An interconnected group of nodes [Olubunmi (2004)]

It is an interconnected assembly of simple processing nodes, whose functionality is based on the animal neuron. Thus, neural networks are an attempt to create systems that work in a similar way to the human brain. These systems are built by using components that act like biological neurons. An artificial neural network can be described as a class of generic filters that stores information in a distributed form.

The processing ability of the network is stored in the inter-unit connection strengths (weights). These weights are obtained by a learning process, which is from a set of training patterns. A system of artificial neural networks may range from a single node to a large collection of nodes in which each node is connected to every other node in the net.

A classification, which is based on the type of learning, architecture and connectivity of the network and learning algorithm, is illustrated in following figure.

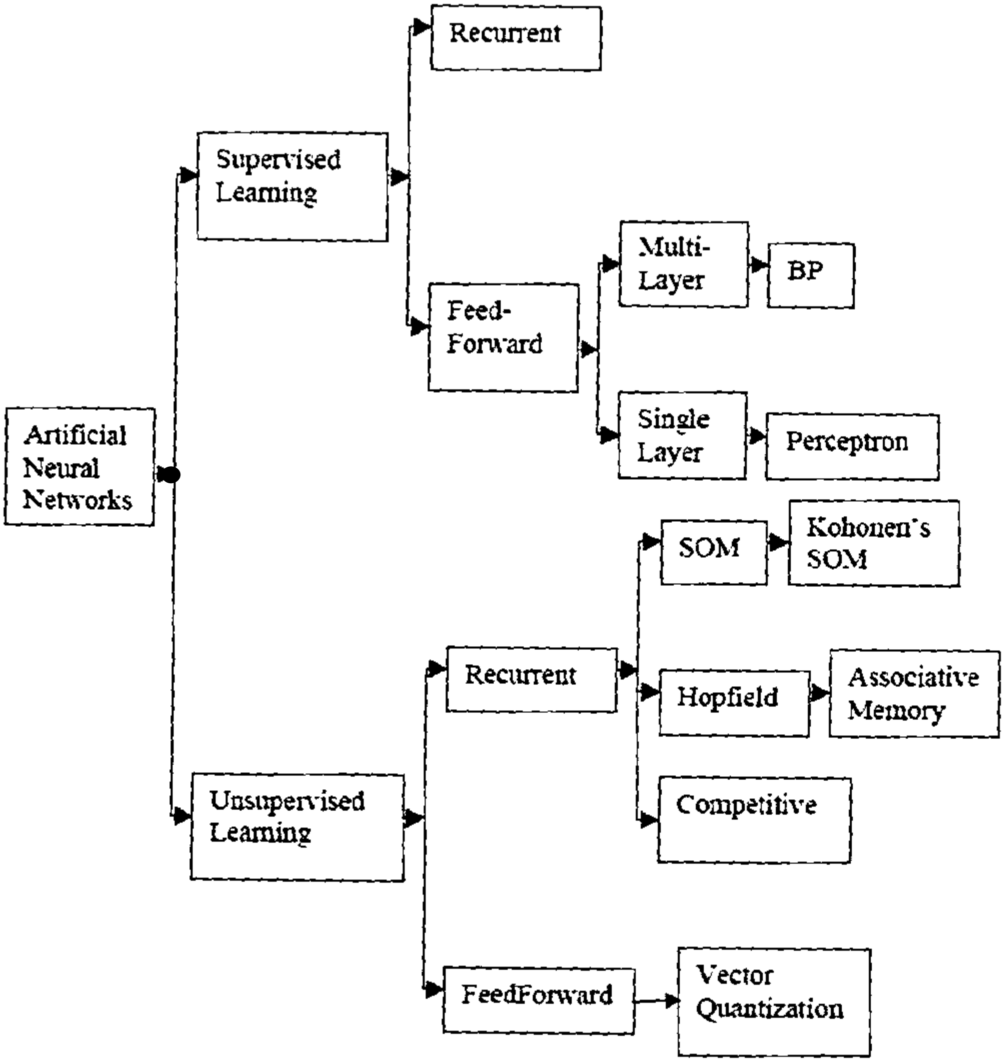

Figure 4: Artificial Neural Network classification model [Olubunmi (2004)]]

The learning process can be *competitive*, meaning that during each training step the particular node is determined that is already closest to the input signal. The node is rewarded by being allowed to adapt even further to the input. in other words: to learn more. If the map receives feedback from the database or outside intervention from the user is necessary, the learning phase is said to be *supervised*. It is called *unsupervised*, if the algorithm learns about the data merely by inspecting it.

2.4.1 Supervised Learning

In supervised learning known class labels help indicate whether the system is performing correctly or not. This information can be used to indicate a desired response, validate the accuracy of the system, or be used to help the system learn to behave correctly. The known class labels can be thought of as supervising the learning process; the term is not meant to imply that you have some sort of interventionist role. ANN Classification is an example of Supervised Learning.

2.4.2 Unsupervised Learning

Consider a machine (or living organism) which receives some sequence of inputs x1, x2, x3, . . ., In supervised learning the machine is also given a sequence of desired outputs y1, y2, . . . , and the goal of the machine is to learn to produce the correct output given a new input. This output could be a class label (in classification) or a real number (in regression).
Finally, in unsupervised learning the machine simply receives inputs x1, x2, . . ., but obtains neither supervised target outputs, nor rewards from its environment. It may seem somewhat mysterious to imagine what the machine could possibly learn given that it doesn't get any feedback from its environment [Zoubin (2004)]. However, it is possible to develop a formal framework for unsupervised learning based on the notion that the machine's goal is to build representations of the input that can be used for decision making, predicting future inputs, efficiently communicating the inputs to another machine, etc. In a sense, unsupervised learning can be thought of as finding patterns in the data above and beyond what would be considered pure unstructured noise. Two very simple classic examples of unsupervised learning are *clustering* and *dimensionality reduction.*

In unsupervised learning the class labels are not presented to the system that is trying to discover the natural classes in a dataset. The purpose of an algorithm for unsupervised learning is to discover significant patterns or clusters in the input data without teacher. Clustering often fails to find known classes because the distinction between the classes can be obscured by the large number of features which are uncorrelated with the classes. A step in ANN classification involves identifying genes which are intimately connected to the known classes. This is called *feature selection* or *feature extraction*. Feature selection and ANN classification together have a use even when prediction of unknown samples is not necessary: They can be used to identify key genes which are involved in whatever processes distinguish the classes.

2.5 The Self-Organizing Map (SOM)

The Self-Organizing Map belongs to the class of unsupervised and competitive learning algorithms. It is a sheet-like neural network, with nodes arranged as a regular, usually two dimensional grids [Jost (2003)]. As explained in the previous section on Neural Networks, we usually think of the node *connections* as being associated with a vector of weights. In the case of Self-Organizing Maps, it is easier to think of each node as being *directly* associated with a weight vector. See Figure 5 below for an illustrative representation of a 4 by 3 map with 3-dimensional weight vectors.

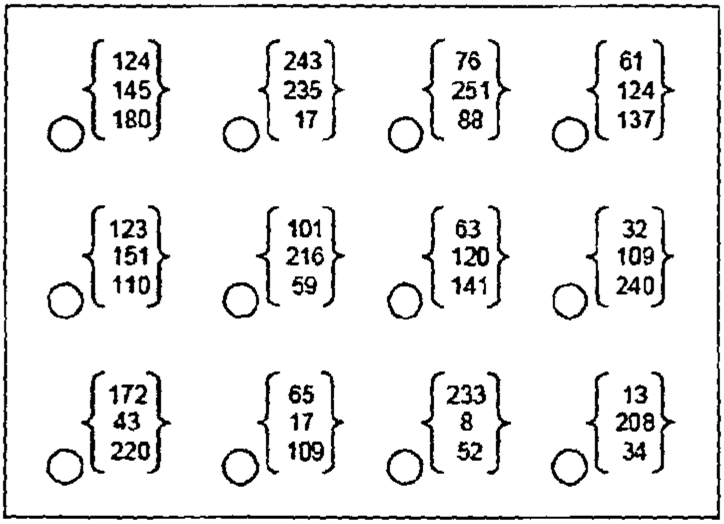

Figure 5: Each map node is associated with a vector of weights [Jost (2003)]

The items in the input data set are assumed to be in a vector format. If *n* is the dimension of the input space, then every node on the map grid holds an n-dimensional vector of weights.

$$m_i = [m_{i1}, m_{i2}, m_{i3}, \ldots\ldots m_{in}] \ldots\ldots\ldots\ldots (2.5.1)$$

The basic principle of the Self-Organizing Map is to adjust these weight vectors until the map represents "a picture" of the input data set. Since the number of map nodes is significantly smaller than the number of items in the dataset, it is needless to say that it is impossible to represent every input item from the data space on the map. Rather, the objective is to achieve a configuration in which the distribution of the data is reflected and the most important metric relationships are preserved. In particular, we are interested in obtaining a correlation between the similarity of items in the dataset and the distance of their most alike representatives on the map. In other words, items that are similar in the input space should map to nearby nodes on the grid.

2.6 The Adaptive Resonance Theory (ART)

The adaptive resonance theory (ART) has been developed to avoid the stability-plasticity dilemma in competitive networks learning. The stability-plasticity dilemma addresses how a learning system can preserve its previously learned knowledge while keeping its ability to learn new patterns. ART architecture models can selforganize in real time producing stable recognition while getting input patterns beyond those originally stored.

ART nets are designed to allow the user to control the degree of similarity of patterns placed on the same cluster. These nets cluster inputs by using unsupervised learning. Input patterns may be presented in any order. Each time a pattern is presented, an appropriate cluster unit is chosen and that cluster's weights are adjusted to let the cluster unit learn the pattern.

As the net is trained, each training pattern may be presented several times. A pattern may be placed on one cluster unit the first time it is presented and then placed on a different cluster when it is presented later (due to changes in the weights for the first cluster if it has learned other patterns in the meantime.) A stable net will not return a pattern to a previous cluster; in other words, a pattern oscillating among different cluster units at different stages of training indicates an unstable net.

The first and most basic architecture is ART1 (Carpenter and Grossberg, 1987a). ART1 is designed to cluster binary input vectors, allowing for great variation in the number of nonzero components and direct user control of the degree of similarity among patterns placed on the same cluster unit. The architecture of an ART1 net consists of two fields of units ----the F1 units(an input processing field) and the F2(cluster) units--- together with a reset unit to control the degree of similarity of patterns placed on the same cluster unit. The F1 and F2 layers are connected by two sets of weighted pathways. In addition, several supplemental units are included in the net to provide for neural control of the learning process.

ART2 (Carpenter and Grossberg, 1987b) is a class of architectures categorizing arbitrary sequences of analog input patterns. ART2 is designed to perform for continuous valued input vectors the same type of tasks as ART1 does for binary valued input vectors. The differences between ART1 and ART2 reflect the modifications needed to accommodate patterns with continuous valued components. The more complex field F1 field of ART2 is necessary because continuous valued input vectors may be arbitrarily close together. The F1 field in ART2 includes a combination of normalization and noise suppression, in addition to the comparison of the bottom-up and top-down signals needed for the reset mechanism. [Fausett (1994)].

2.7 The Learning Vector Quantization (LVQ)

Learning Vector Quantization (LVQ) is a pattern classification method in which each output unit represents a particular class or category. The weight vector for an output unit is often reoffered to as a reference (or codebook) vector for the class that the unit represents. During training, the output units are positioned (by adjusting their weights) to approximate the decision surfaces of the theoretical Bayes classifiers. It is assumed that a set of training patterns with known classification is provided, along with an initial distribution of reference vectors. After training, an LVQ net classifies an input vector by assigning it to the same class as the output unit has its weight vector closest to the input vector.

There are a number of improved LVQ algorithms, known as LVQ 2, LVQ2.1 and LVQ3. In the original LVQ algorithm, only the reference vector that is closest to the input vector is updated. The direction it is moved depends on whether the winning reference vector belongs to the same class as the input vector. In the improved algorithms, two vectors (the winner and runner-up) learn if several conditions are satisfied. The idea is that if the input is approximately the same distance from both the winner and the runner-up, then each of them should learn [Fausett (1994)].

Chapter 3

REQUIRMENT ANALYSIS

3. Requirement Analysis

This chapter aims to specify what exactly the system is expected to deliver.

3.1 Introduction

The telecommunications companies generate and store a tremendous amount of data. This data include:-

- Call detail data, which describes the calls that traverse the telecommunication networks,
- Network data, which describes the state of the hardware and software components in the network, and
- Customer data, which describes the telecommunication customers.

The amount of data is so great that manual analysis of the data is difficult, if not impossible. Insight into information and knowledge derived from these databases can give operators a competitive edge in terms of customer care and retention, marketing and fraud detection.

In software engineering, requirement analysis encompasses those tasks that go into determining the needs or conditions to meet for a new or altered product, taking account of the possibly conflicting requirements of the various stakeholders, such as beneficiaries or users. Systematic requirements analysis is also known as *requirements engineering*. It is sometimes referred to loosely by names such as *requirements gathering*, *requirements capture*, or *requirements specification*. The term *requirements analysis* can also be applied specifically to the analysis proper (as opposed to elicitation or documentation of the requirements, for instance). Requirements analysis is critical to the success of a development project.

Requirements must be actionable, measurable, testable, related to identified business needs or opportunities, and defined to a level of detail sufficient for system design.

3.2 Problem Statement

Although SOM has been used for Trend Analysis of different types of data including Telecom but LVQ and ART has not been used as a basis where unsupervised learning in a dynamic model fashion is applied to Telecom data for trend analysis. Also comparative study of these unsupervised neural network algorithms has not yet been carried out. Hence the effectiveness and comparative analysis of these different techniques has not yet been explored in detail. Therefore, until and unless we implement these techniques on the data, we will be unable to predict accurately the selection and application of individual algorithms in different cases on telecom data.

The telecommunications industry generates and stores a tremendous amount of data. These data include call detail data, which describes the calls that traverse the telecommunication networks, network data, which describes the state of the hardware and software components in the network, and customer data, which describes the telecommunication customers. The amount of data is so great that manual analysis of the data is difficult, if not impossible. The need to handle such large volumes of data led to the development of knowledge-based expert systems [Jacome and Lanca (1989); Kim et al. (1992b)] These automated systems performed important functions such as identifying fraudulent phone calls and identifying network faults. The problem with this approach is that it is time consuming to obtain the knowledge from human experts (the "knowledge acquisition bottleneck") and, in many cases; the experts do not have the requisite knowledge. The advent of data mining technology promised solutions to these problems.

Unsupervised neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyze. Unsupervised learning is very useful and remarkable technique when dealing with large volumes of raw data with little or no knowledge of the inter relation between the various fields in a vector. This network can create its own organization or representation of the information it receives during learning time.

An unsupervised learning algorithm can analyze and cluster call patterns for each subscriber in order to facilitate the trend analysis. Our study provides a comparative analysis and application of Self-Organizing Maps (SOM), Learning Vector Quantization (LVQ) and Adaptive Resonance Theory (ART) neural networks algorithms to user call data records in order to conduct a descriptive trend analysis in the hidden data.

The objective is to build a generic system based on the unsupervised neural network algorithms for the trend analysis of Telecom data. The system must be capable of generating and visualizing graphs, thus displaying the trend and cluster relationships hidden in the data.

3.3 Problem Scenarios

In the telecommunications industry, it is often useful to profile customers based on their patterns of phone usage, which can be extracted from the call detail data. These customer profiles can then be used for marketing purposes, or to better understand the customer, which in turn may lead to better forecasting models. In order to effectively mine the call detail data, it must be summarized to the customer level. Telecom operators of today have access to a lot of data regarding their subscribers. This includes among many other things subscription data, service usage, and location data. This data can be analyzed using different data mining techniques to better understand the end-user behavior and the end-user needs. The gained knowledge about the end-users can then be used to improve existing services, target specific user groups with service offerings or to simply share this knowledge with 3rd party service providers to improve the services they are offering.

The telecommunications companies generate and store a tremendous amount of data. This data include:-

- Call detail data, which describes the calls that traverse the telecommunication networks,
- Network data, which describes the state of the hardware and software components in the network, and
- Customer data, which describes the telecommunication customers.

The amount of data is so great that manual analysis of the data is difficult, if not impossible. The need to handle such large volumes of data led to the development of knowledge-based expert systems. Insight into information and knowledge derived from these databases can give operators a competitive edge in terms of customer care and retention, marketing and fraud detection.

One of the strategies for marketing campaign analyzes the patterns or trends of the most / least revenue generating customers. Although the customer and their amount of revenue changes over time, their common characteristics and attributes are reflected in the customer and call detail data. Over the period of time, customer revenue is collected from its call detail data and other characteristics are available in customer data. Further analysis is thus, required to be able to analyze trend in customer data w.r.t revenue.
Another strategy for marketing campaign analyzes the patterns or trends of the network usage with time. Different customers use network in different timings as per their requirement. For better usage of network and load balancing, it is necessary to analyze the network usage with respect to time. Further analysis is thus, required to be able to analyze trend in network usage data w.r.t time. An unsupervised learning algorithm can analyze and cluster call detail records for each subscriber in order to analyze trends in customer revenue generating process. Also this can analyze and cluster network usage data for time in order to analyze trends in network usage process.

Specifically, we are taking two problem scenarios:

1. How to find the pattern or trend in the collected information of the most or least revenue generating customers over a time period?

2. How to find the pattern or trend in the collected information of the network usage with time?

Chapter 4

SYSTEM DESIGN

4. System Design

During Literature Survey, we have seen that although some work has been done on the Trend Analysis of Telecom data using SOM but LVQ and ART have not been applied to Telecom data for trend analysis. In addition, comparative study of these unsupervised neural network algorithms has not yet been carried out.

Hence we have proposed a solution in MATLAB by implementing SOM, LVQ and ART algorithms. The clusters will be generated on the basis of user revenues and network usage. Each algorithm will take input and then categorize into different clusters based on the above three NN techniques. These clusters will form the basis for data and trend analysis. The graphs generated will portray the trends hidden in the data within a cluster as well as across the clusters. Also we will provide a comparative analysis of the outputs/results generated by Self-Organizing Maps (SOM), Learning Vector Quantization (LVQ) and Adaptive Resonance Theory (ART) unsupervised neural networks algorithms. Important characteristics and advantages of one algorithm over the other will be analyzed.

In this chapter we will discuss the System Design methodology.

4.1 Introduction

Systems design is the process or art of defining the architecture, components, modules, interfaces, and data for a system to satisfy specified requirements. The data mining process must be reliable and repeatable by people with little data mining skills. We have used the CRISP-DM (Cross Industry Standard Process for Data Mining) methodology for the system design.

4.2 Reference Architecture

The CRISP-DM is an industry- and tool-neutral data mining process model [CRISP-DM (2007)]. Starting from the embryonic knowledge discovery processes used in early data mining projects and responding directly to user requirements, this project defined and validated a data mining process that is applicable in diverse industry sectors. This methodology makes large data mining projects faster, cheaper, more reliable and more manageable. Even small scale data mining investigations benefit from using CRISP-DM.

The life cycle of a data mining project consists of six phases. The sequence of the phases is not strict. Moving back and forth between different phases is always required. It depends on the outcome of each phase which phase, or which particular task of a phase, that has to be performed next. The arrows indicate the most important and frequent dependencies between phases.

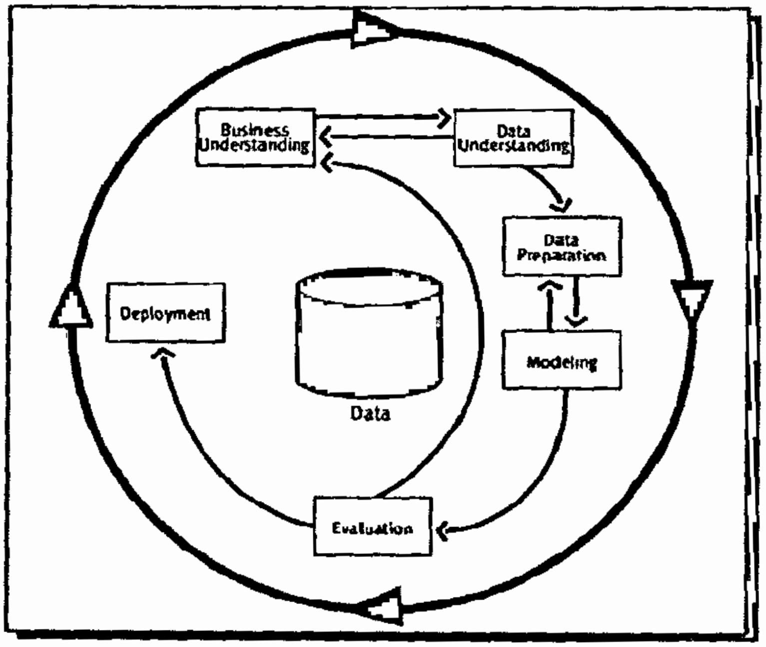

Figure 6: CRISP-DM (Cross Industry Standard Process for Data Mining) [CRISP-DM (2007)]

The outer circle in the figure symbolizes the cyclic nature of data mining itself. A data mining process continues after a solution has been deployed. The lessons learned during the process can trigger new, often more focused business questions. Subsequent data mining processes will benefit from the experiences of previous ones.

Below follows a brief outline of the phases:

Business Understanding

This initial phase focuses on understanding the project objectives and requirements from a business perspective, and then converting this knowledge into a data mining problem definition, and a preliminary plan designed to achieve the objectives.

Data Understanding

The data understanding phase starts with an initial data collection and proceeds with activities in order to get familiar with the data, to identify data quality problems, to discover first insights into the data, or to detect interesting subsets to form hypotheses for hidden information.

Data Preparation

The data preparation phase covers all activities to construct the final dataset (data that will be fed into the modeling tool(s)) from the initial raw data. Data preparation tasks are likely to be performed multiple times, and not in any prescribed order. Tasks include table, record, and attribute selection as well as transformation and cleaning of data for modeling tools.

Modeling

In this phase, various modeling techniques are selected and applied, and their parameters are calibrated to optimal values. Typically, there are several techniques for the same data mining problem type. Some techniques have specific requirements on the form of data. Therefore, stepping back to the data preparation phase is often needed.

Evaluation

At this stage in the project we have built a model (or models) that appears to have high quality, from a data analysis perspective. Before proceeding to final deployment of the model, it is important to more thoroughly evaluate the model, and review the steps executed to construct the model, to be certain it properly achieves the business objectives. A key objective is to determine if there is some important business issue that has not been sufficiently considered. At the end of this phase, a decision on the use of the data mining results should be reached.

Deployment

Creation of the model is generally not the end of the project. Even if the purpose of the model is to increase knowledge of the data, the knowledge gained will need to be organized and presented in a way that the customer can use it. Depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process. In many cases it will be the customer, not the data analyst, who will carry out the deployment steps. However, even if the analyst will not carry out the deployment effort it is important for the customer to understand up front what actions will need to be carried out in order to actually make use of the created models?

4.3 Design Methodology

Figure 8 illustrates design methodology.

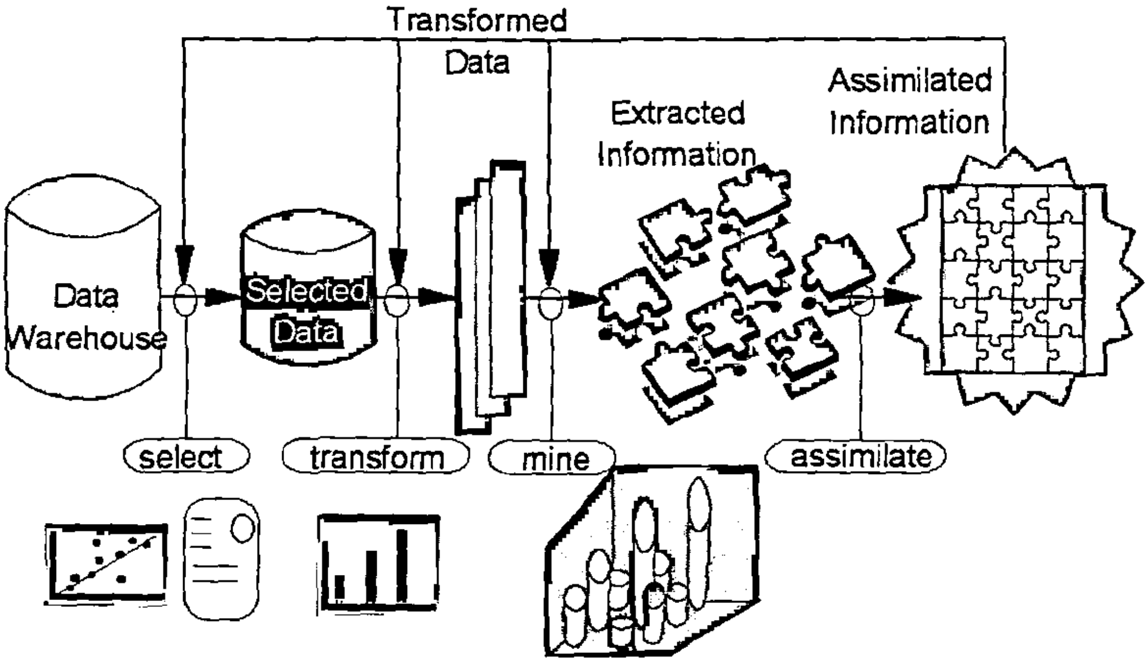

Figure 7: Data Mining Design Methodology

4.4 Algorithms Implemented.

We have implemented three algorithms namely Self-Organizing Maps (SOM), Learning Vector Quantization (LVQ) and Adaptive Resonance Theory (ART).

4.4.1 The Self-Organizing Map (SOM)

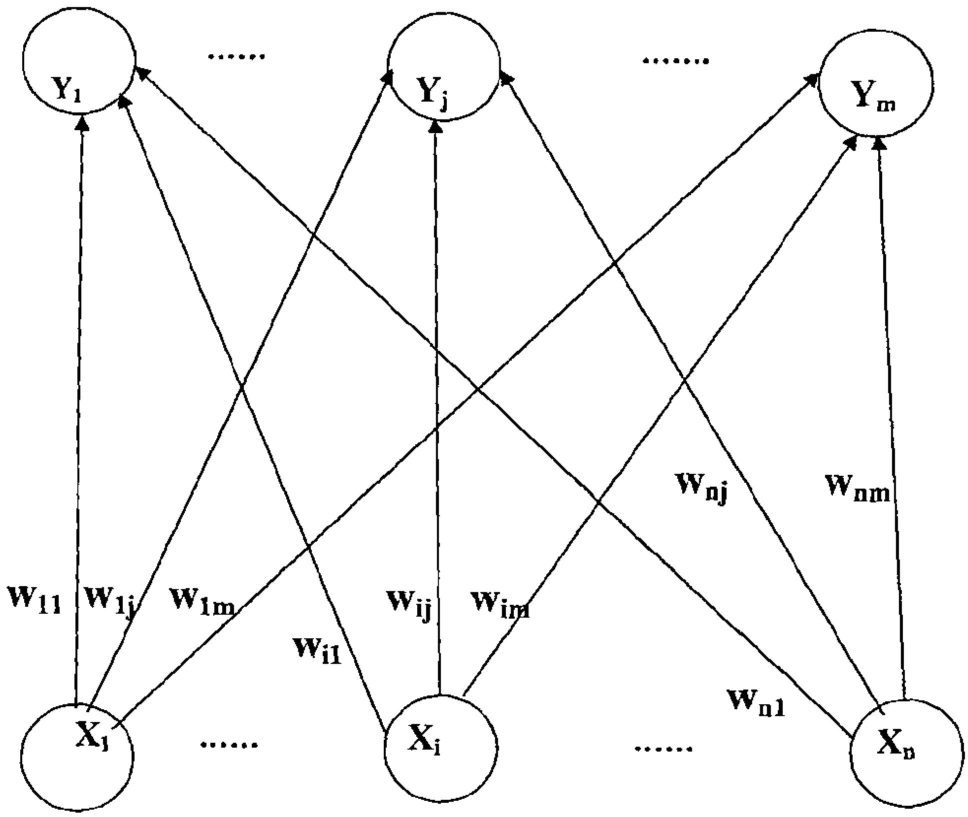

Figure 8a: Typical SOM architecture [Fausett (1994)]

ALGORITHM

Step 0. Initilize weights w_{ij}. Set topological neighborhood parameters. Set learning rate parameters

Step 1. While stopping condition is false, do Steps 2-8.

Step 2. For each input vector x, do Steps 3-5.

Step 3. For each j,compute:

$$D(j) = \sum_i (w_{ij} - x_i)^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4.4.1.1).$$

Step 4. Find index J such that D(J) is a minimum.

Step 5. For all units j within a specified neighborhood of J, and for all i:

$$w_{ij}(new) = w_{ij}(old) + \alpha[x_i - w_{ij}(old)] \text{ ... ... (4.4.1.2)}$$

Step 6. Update learning rate.

Step 7. Reduce radius of topological neighborhood at specified times.

Step 8. Test stopping condition.

The choice of parameters

x training vector $(x_1 \dots \dots . x_i \dots \dots x_n$

T correct category or class for the training vector.

w_j weight vector for jth output unit $(w_{1j} \dots \dots . w_{ij} \dots \dots w_{nj}$

C_j category or class represented by jth output unit.

$\|x - w_j\|$ Euclidean distance between input vector and (weight vector for) jth output unit.

α learning rate.

The learning rate α is a slowly decreasing function of time. [Kohonen (1989a, p.133)] indicates that a linearly decreasing function is satisfactory for practical computations. The radius of the neighborhood around a cluster unit decreases as the clustering process progresses.

The formation of the map occurs in two phases: the initial formation of the correct order and the final convergence. The second phase takes much longer than the first and requires a small value for the learning rate.

Random values may be assigned for the initial weights. If some information is available concerning the distribution of clusters that might be appropriate for a particular problem, the initial weights can be taken to reflect that prior knowledge.

4.4.2 The Adaptive Resonance Theory (ART)

We have selected ART2 algorithm for continued-valued inputs

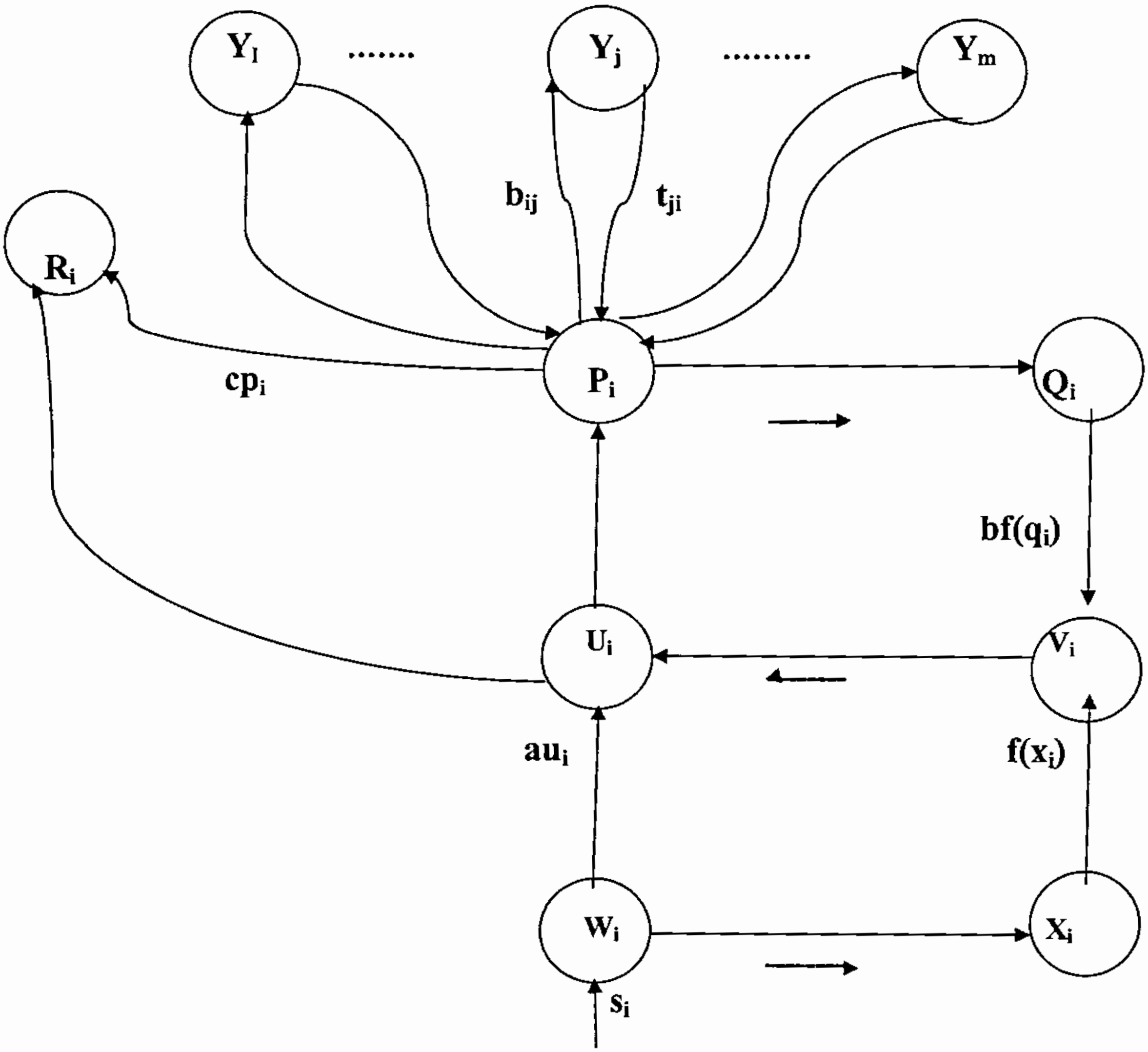

Figure 8b: Typical ART2 architecture [Fausett (1994)]

This section provides a description for the training algorithm for ART2 (for both fast and slow learning) and a step by step statement of the algorithm.

ALGORITHM

Step 0. Initialize parameters:

$a, b, \theta, c, d, e, \alpha, \rho.$

Step 1. Do Step 2-12 N_EP times.
(Perform the specified number of epochs of training.)

Step 2. For each input vector s, do Steps 3-11.

Step 3. Update F1 unit activations:

$$u_i = 0 \dots\dots\dots\dots\dots\dots\dots (4.4.2.1)$$

$$w_i = s_i \dots\dots\dots\dots\dots\dots\dots (4.4.2.2)$$

$$p_i = 0 \dots\dots\dots\dots\dots\dots\dots (4.4.2.3)$$

$$x_i = \frac{s_i}{e+\|s\|} \dots\dots\dots\dots\dots\dots (4.4.2.4)$$

$$q_i = 0 \dots\dots\dots\dots\dots\dots\dots (4.4.2.5)$$

$$v_i = f(x_i) \dots\dots\dots\dots\dots\dots (4.4.2.6)$$

Update F_1 unit activations again:

$$u_i = \frac{v_i}{e+\|v\|} \dots\dots\dots\dots\dots\dots (4.4.2.7)$$

$$w_i = s_i + au_i \dots\dots\dots\dots\dots (4.4.2.8)$$

$$p_i = u_i \dots\dots\dots\dots\dots\dots\dots (4.4.2.9)$$

$$x_i = \frac{w_i}{e+\|w\|} \dots\dots\dots\dots\dots\dots (4.4.2.10)$$

$$q_i = \frac{p_i}{e+\|p\|} \dots\dots\dots\dots\dots\dots (4.4.2.11)$$

$$v_i = f(x_i) + bf(q_i) \dots\dots (4.4.2.12)$$

a, b fixed weights in the F1 layer; sample values are a = 10, b = 10. Setting either a = 0 or b = 0 produces instability in the net; other than that, the net is not particularly sensitive to the values chosen.

c fixed weight used in testing for reset; a sample value is c = .1. A small c gives a larger effective range of the vigilance parameter

d activation of winning F_2 unit; a sample value is d =.9. Note that c and d must be chosen to satisfy the inequality

$$\frac{cd}{1-d} \leq 1$$

n (in order to prevent a reset from occurring during a learning trial). The ratio should be chosen close to 1 to achieve a larger effective range for vigilance.

e a small parameter introduced to prevent division by zero when the norm of a vector is zero. This value prevents the normalization to unity from being exact. A value of zero is typically used in the hand computations and derivations that follow and may be used in the algorithm if the normalization step is skipped when the vector is zero.

θ noise suppression parameter, a sample value is $\theta = {}^{1}/_{\sqrt{n}}$. The sample value may be larger than desired in some applications. Components of the normalized input vector (and other vectors in the F_1 loop) that are less than this value are set to zero.

α learning rate. A smaller value will slow the learning in either the fast or the slow learning mode. However, a smaller value will ensure that the weights (as well as the placement of patterns on clusters) eventually reach equilibrium in the slow learning mode).

ρ vigilance parameter. Along with the initial bottom-up weights, this parameter determines how many clusters will be formed. Although, the theoretically, values from 0 to 1 are allowed, only values between approximately 0.7 to 1 perform any useful role in controlling the number of clusters. (Any value less than 0.7 will have the same effect as setting ρ to zero.) Some choices of values for c and d will restrict the effective range of values for ρ even further.

4.4.3 The Learning Vector Quantization (LVQ)

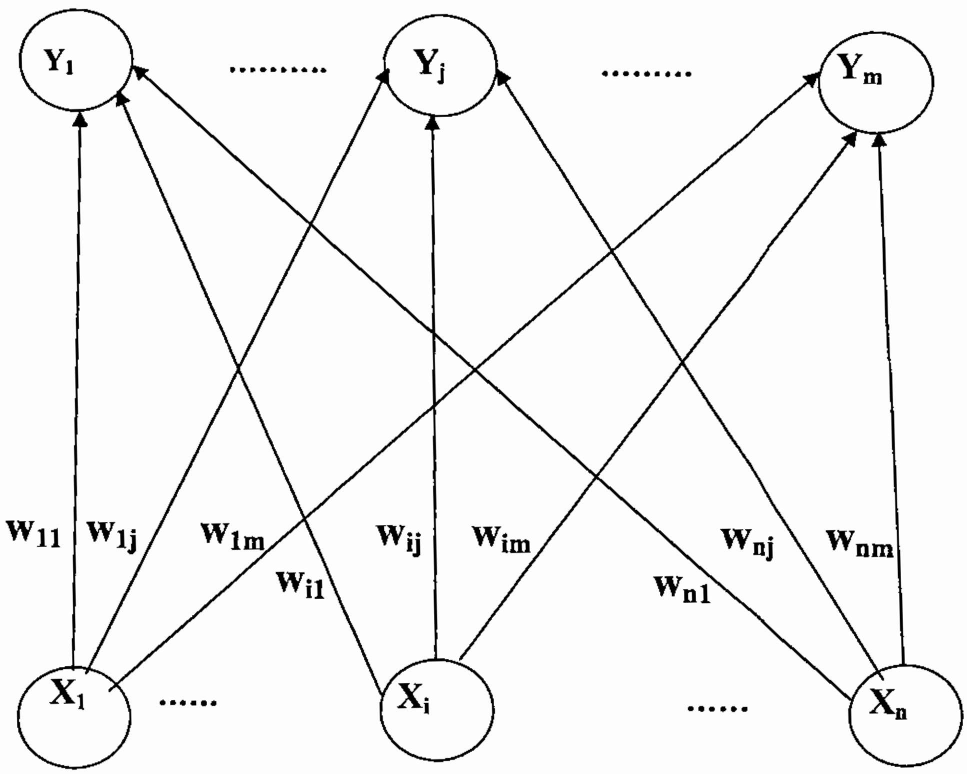

Figure 8c: Typical ART2 architecture [Fausett (1994)]

ALGORITHM

Step 0. Initialize reference vectors. Initialize learning rate, $\alpha(0)$.

Step 1. While stopping condition is false, do Steps 2 – 6.

Step 2. For each training input vector x, do Steps 3 – 4.

Step 3. Find J so that $\|x - w_J\|$ is a minimum.

Step 4. Update w_J as follows

If $T = C_J$, then

$$w_J(new) = w_J(old) + \alpha[x - w_J(old)] \quad \ldots\ldots\ldots\ldots\ldots\ldots (4.4.3.1)$$

Chapter 5

IMPLEMENTATION & RESULTS

5. Implementation and Results

During Literature Survey, we have seen that although some work has been done on the Trend Analysis of Telecom data using SOM but LVQ and ART have not been applied to Telecom data for trend analysis. In addition, comparative study of these unsupervised neural network algorithms has not yet been carried out.

Hence we have proposed a solution in MATLAB by implementing SOM, LVQ and ART algorithms. The clusters will be generated on the basis of user revenues and network usage. Each algorithm will take input and then categorize into different clusters based on the above three NN techniques. These clusters will form the basis for data and trend analysis. The graphs generated will portray the trends hidden in the data within a cluster as well as across the clusters. Also we will provide a comparative analysis of the outputs/results generated by Self-Organizing Maps (SOM), Learning Vector Quantization (LVQ) and Adaptive Resonance Theory (ART) unsupervised neural networks algorithms. Important characteristics and advantages of one algorithm over the other will be analyzed.

We have used the CRISP-DM (Cross Industry Standard Process for Data Mining) methodology for the system design. The initial phase focuses on understanding the project objectives and requirements from a business perspective, and then converting this knowledge into a data mining problem definition, and a preliminary plan designed to achieve the objectives. The data understanding phase starts with an initial data collection and proceeds with activities in order to get familiar with the data. The data preparation phase covers all activities to construct the final dataset (data that will be fed into the modeling tool(s)) from the initial raw data. Then, various modeling techniques (SOM, LVQ and ART) are selected and applied, and their parameters are calibrated to optimal values.

The implementation process is mainly consisted of three phases:

1. Data Selection and Preparation
2. Model Building
3. Evaluation

5.1 Data Selection and Preparation

The first phase is data selection and preparation, where the data and variables of interest are selected, transformed and normalized from the data warehouse of the enterprise. The data used in the project was from a Pakistani GSM operator.

Every time a call is placed on a telecommunications network, descriptive information about the call is saved as a *call detail* record. The number of call detail records that are generated and stored is huge. Call detail records include sufficient information to describe the important characteristics of each call. At a minimum, each call detail record will include the originating and terminating phone numbers, the date and time of the call, and the cost and duration of the call.

Telecommunication companies, like other large businesses, may have millions of customers / subscribers. By necessity this means maintaining a database of information on these customers. This information include name and address information and may include other information such as product and service plan and contract information, credit score and other historical information.

We extracted only Mobile Originating Calls for two weeks and only three attributes of subscriber. Call detail records are not used directly for data mining, since the goal of data mining applications is to extract knowledge at the customer level, not at the level of individual phone calls. Thus, the call detail records associated with a customer must be summarized into a single record that describes the customer's calling behavior.

The finalized selected data after normalization contained the following fields.

1.	Subscriber Number	(MSISDN)
2.	City for the location of the subscriber	(CITY)
3.	Product / Service used by Subscriber	(PACKAGE)
4.	Date and time the call was made	(CallDateTime)
5.	Cost of Call	(Revenue)
6.	Duration of Call	(Minutes)

The subscriber number was not used in training but to identify the trends and patterns of each subscriber.

5.2 Model Building

In this phase, the actual *data mining* begins. Unsupervised neural network technique is selected for model building. The following three algorithms are selected and implemented using MATLAB.

1. Self Organizing Map (SOM)
2. Learning Vector Quantization (LVQ)
3. Adaptive Resonance Theory (ART)

5.2.1 Self Organizing Map (SOM)

Kohonen Self Organizing Map is used and the empirical values selected for different parameters and techniques are as follows:

1. Initial Radius: R = 0
2. Learning Rate: $\alpha = 0.6$

3. Total Number of Iterations = 100
4. Weight Initialization = Random
5. Topology = Linear

5.2.2 Learning Vector Quantization (LVQ)

1. Weight Initialization Method = SOM
2. Number of Training Vectors = 100
3. Learning Rate: $\alpha = 0.1$
4. Total Number of Learning Iterations = 100
5. Topology = Linear

5.2.3 Adaptive Resonance Theory (ART)

ART2 algorithm is an extension of ART1. The main difference between these two is that ART2 is designed to perform for continuous-valued input vectors where as the ART1 is only for binary input vector. The empirical values selected for different parameters and techniques are as follows:

1. Fixed Weights in F1 layer: a = b = 10
2. Fixed weight used in testing for reset: c = 0.1
3. Activation of winning F2 unit: d = 0.9
4. Noise suppression parameter: $\theta = 0.3$
5. Learning Rate: $\alpha = 0.39$
6. Vigilance parameter: $\rho = 0.9$
7. Learning Technique = Fast Learning

5.3 Evaluation

5.3.1 Clustering Evaluation

In this phase model is evaluated on input data and resultant clusters are analyzed with input data.

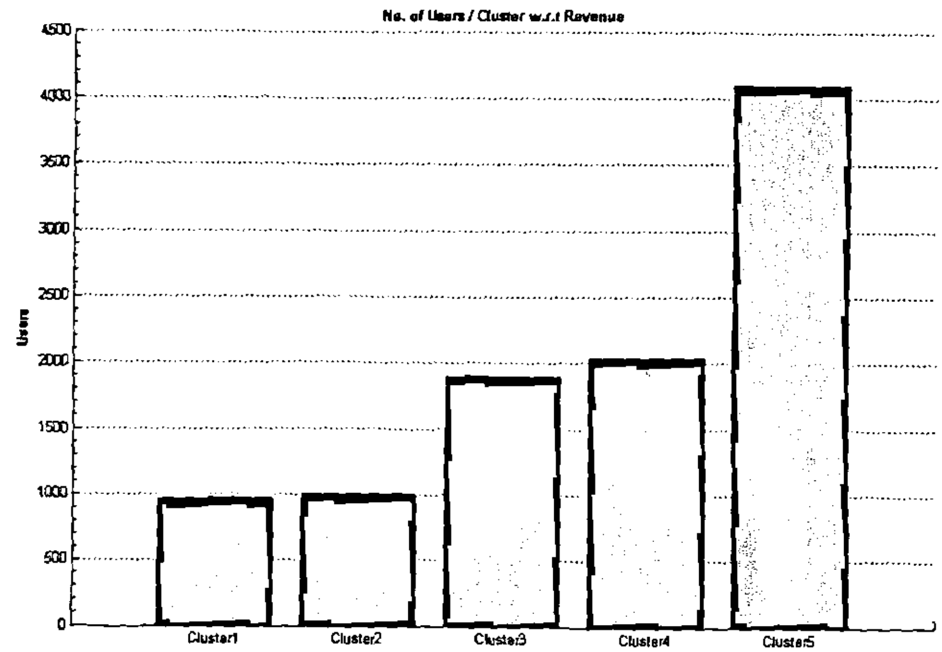

Figure 11: LVQ based Clusters according to Revenue

Figure 11 shows the categorization formed by LVQ algorithm. LVQ has categorized the input data into 5 clusters of subscribers based on their 15 days revenue. The data is categorized into five clusters i-e Subscribers with Most High Revenue, High Revenue, Average Revenue, Low Revenue and Least Revenue.

The only difference between SOM and LVQ segmentation is the order of categorization of subscribers in different clusters is opposite to the order of clusters of SOM.

Also it shows that the 10,000 subscribers are segmented as follows:

Cluster	No. of Subscribers	Cluster Explanation
1	800	Most High Revenue producing customers
2	1,000	High Revenue producing customers
3	1,900	Average Revenue producing customers
4	2,200	Low Revenue producing customers
5	4,100	Least Revenue producing customers

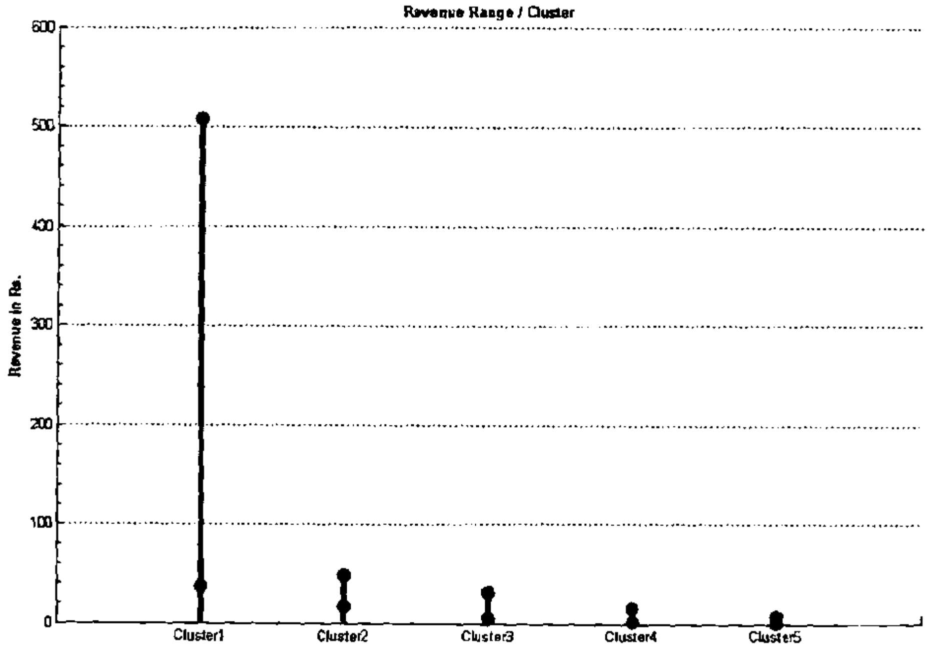

Figure 12: LVQ based Cluster Ranges according to Revenue

The range of revenues of subscribers in different clusters by LVQ is shown in Figure 12.

Cluster	Range of Revenue	Cluster Explanation
1	40~500	Most High Revenue of customers
2	25~60	High Revenue of customers
3	>20~40	Average Revenue of customers
4	>10~20	Low Revenue of customers
5	0~10	Least Revenue of customers

The analysis is same as of SOM which shows that Most High Revenue generating customers are very short in numbers but they are most valuable customers to the business and a special care is required to retain these customers.

Similarly, the number of Least/Low Revenue generating customers is very high hence there should be marketing promotions to get more and more revenue from these subscribers.

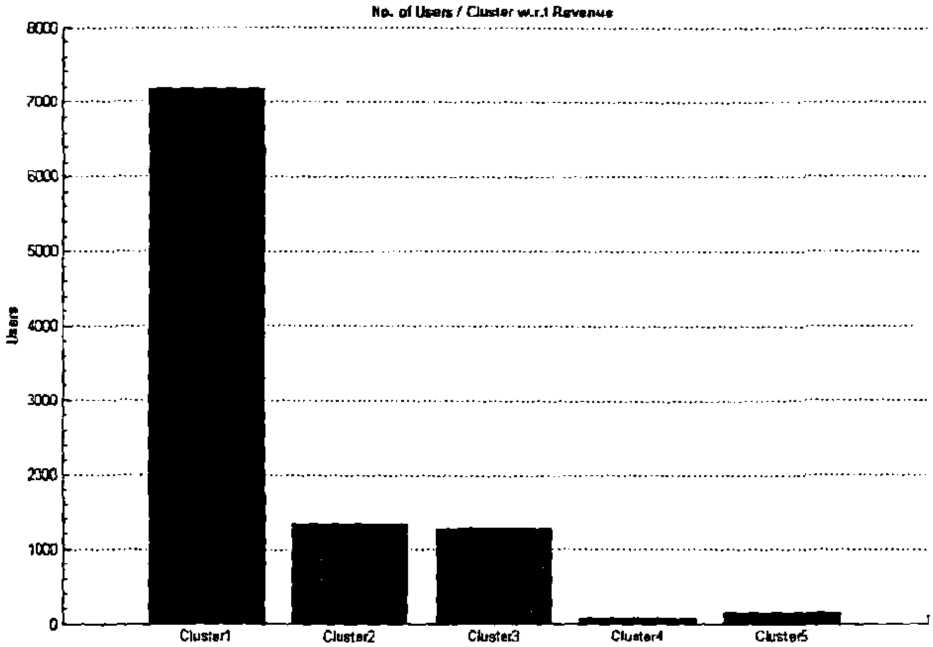

Figure 13: ART based Clusters according to Revenue

Figure 13 shows the categorization formed by ART algorithm. The clusters formed by ART are totally different to the clusters formed by SOM and LVQ. The reason of this difference is that ART segments data based on all the given components of input vector. So a slight variation in second component of input data will be place in different cluster/segment by ART algorithm.
ART categorizes the input data into 5 clusters of subscribers based on their 15 days revenue. The data is categorized into five clusters on the basis of different input vectors not ranges of revenue data. Also it shows that the 10,000 subscribers are segmented as follows:

Cluster	No. of Subscribers	Cluster Explanation
1	7,100	Most High Revenue
2	1,300	High Revenue
3	1,200	Average Revenue
4	300	Least Revenue
5	100	Low Revenue

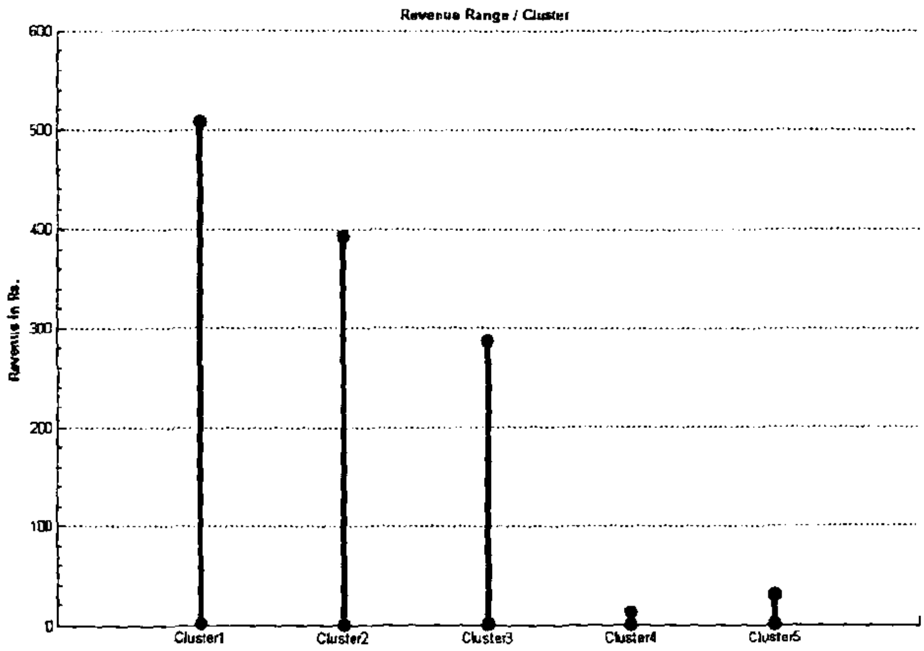

Figure 14: ART based Cluster Ranges according to Revenue

The range of revenues of subscribers in different clusters is shown in Figure 14.

Cluster	Range of Revenue	Cluster Explanation
1	0~500	Most High Revenue of customers
2	0~380	High Revenue of customers
3	0~280	Average Revenue of customers
4	0~10	Least Revenue of customers
5	0~40	Low High Revenue of customers

The analysis obtained by ART algorithm is not satisfied with given data because of input measures selected for analysis with this algorithm.

5.3.1.2 Clustering of Time of call Data

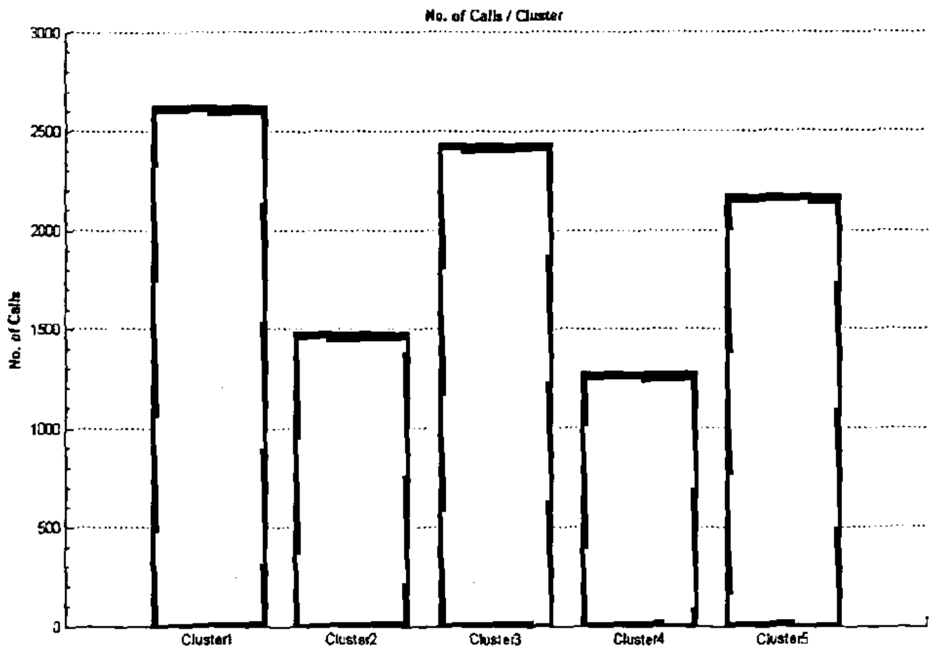

Figure 15: SOM based Clusters for Time of Call

Figure 15 shows the categorization formed by SOM algorithm. SOM categorizes the input data into 5 clusters of Subscribers with respect to their Call Timings based on their 15 days CDR data. The data is categorized into five clusters i-e Subscribers with Call Time 7 P.M. to 11 P.M, 5 P.M. to 6 P.M, 1 P.M. to 4 P.M, 11 A.M. to 12 P.M, 1 A.M. to 10 A.M. Also it shows that the 10,000 subscribers are segmented as follows:

Cluster	No. of Subscribers	Cluster Explanation
1	2,600	Most Busy Network w.r.t Traffic
2	1,500	Average Network w.r.t Traffic
3	2,400	Busy of Network w.r.t Traffic
4	1,300	Not Busy Network w.r.t Traffic
5	2,200	Busy Network w.r.t Traffic

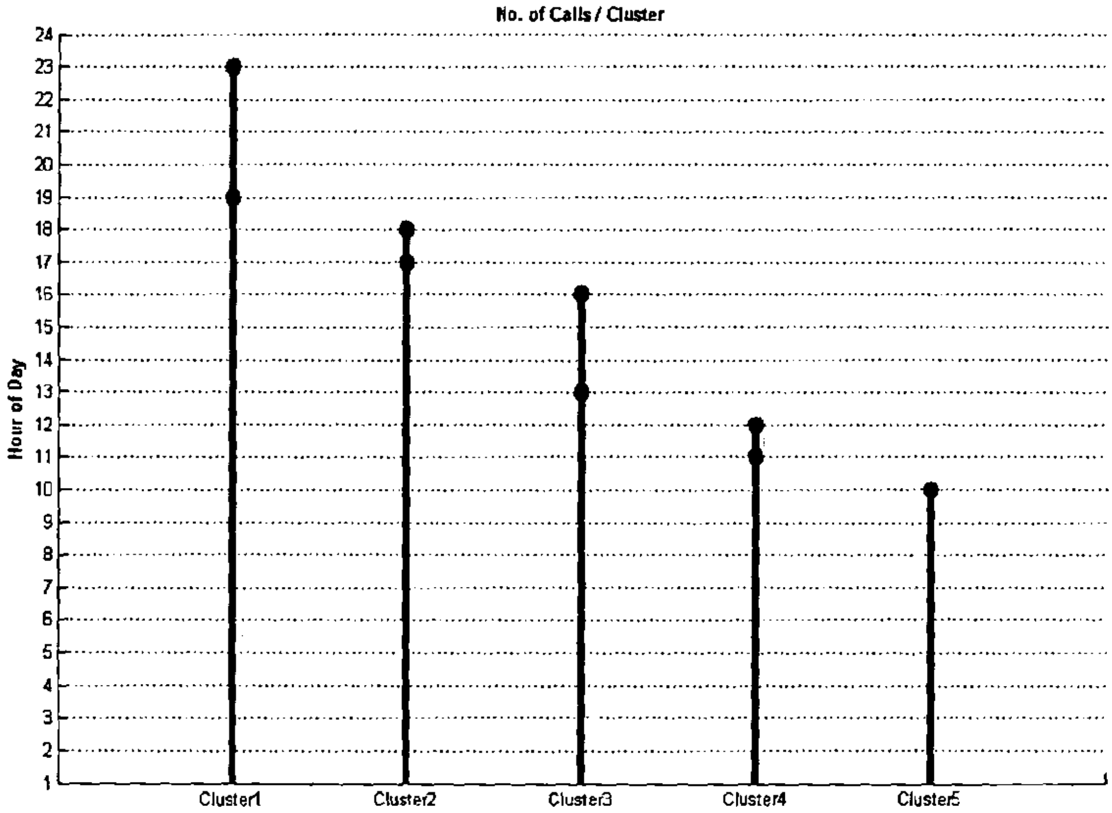

Figure 16: SOM based Cluster Ranges for Time of Call

The range of Hours w.r.t Call Time in different clusters by SOM is shown in Figure 16.

Cluster	Range of Hours	Cluster Explanation
1	7 P.M. to 11 P.M	Most Busy Network w.r.t Traffic
2	5 P.M. to 6 P.M	Average Network w.r.t Traffic
3	1 P.M. to 4 P.M	Busy Network w.r.t Traffic
4	11 A.M. to 12 P.M	Not Busy Network w.r.t Traffic
5	12 A.M. to 10 A.M	Busy Network w.r.t Traffic

The analysis shows that Network is Most Busy w.r.t Traffic from 7 P.M to 11 P.M so special care is required during these Peak Hours for Network maintenance. Also the call rate should be adjusted during these Peak Hours of Traffic. Similarly, the Non Peak Hours for Network Traffic are 11 A.M to 12 P.M. Marketing campaigns should address these Hours and different rates are offered for these Off Peak hours for better utilization of network and hence to produce more revenue for business.

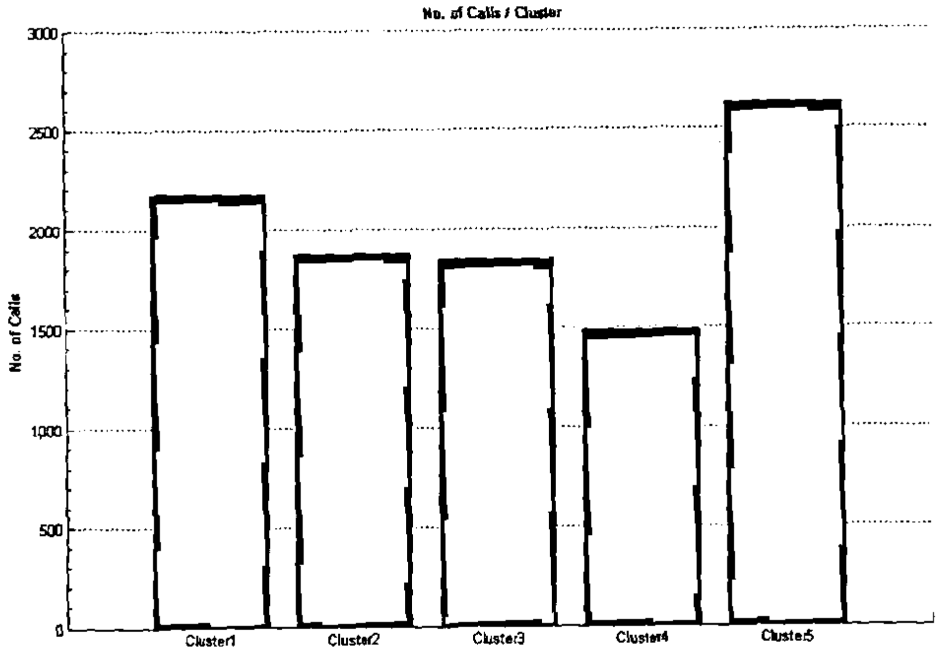

Figure 17: LVQ based Clusters for Time of Call

Figure 17 shows the categorization formed by LVQ algorithm. This clustering is some what different with the SOM clustering. LVQ categorizes the input data into 5 clusters of Subscribers with respect to their Call Timings based on their 15 days CDR data. The data is categorized into five clusters i-e Subscribers with Call Time 12 A.M. to 10 A.M, 11 A.M. to 1 P.M, 2 P.M. to 4 P.M, 5 P.M. to 6 P.M, and 7 P.M. to 11 P.M. Also it shows that the 10,000 subscribers are segmented as follows:

Cluster	No. of Subscribers	Cluster Explanation
1	2,200	More Busy Network w.r.t Traffic
2	1,900	Busy Network w.r.t Traffic
3	1,800	Busy of Network w.r.t Traffic
4	1,500	Not Busy Network w.r.t Traffic
5	2,600	Most Busy Network w.r.t Traffic

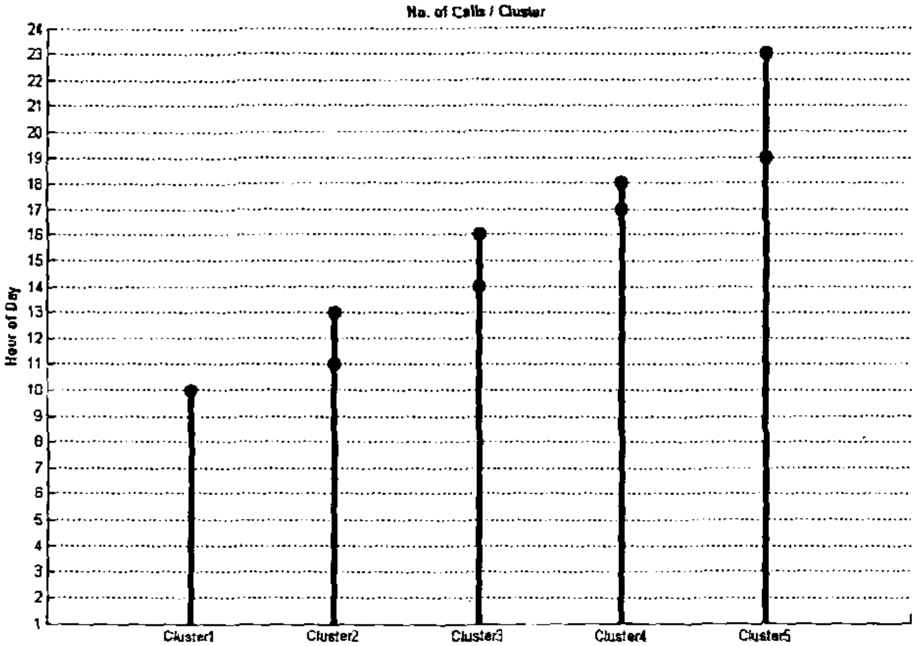

Figure 18: LVQ based Cluster Ranges for Time of Call

The range of Hours w.r.t Call Time in different clusters by LVQ is shown in Figure 18.

Cluster	Range of Hours	Cluster Explanation
1	12 A.M. to 10 A.M	Busy Network w.r.t Traffic
2	11 A.M. to 1 P.M	Not Busy Network w.r.t Traffic
3	2 P.M. to 4 P.M	Busy Network w.r.t Traffic
4	5 P.M. to 6 P.M	Average Network w.r.t Traffic
5	7 P.M. to 11 P.M	Most Busy Network w.r.t Traffic

The analysis shows that Network is Most Busy w.r.t Traffic from 7 P.M to 11 P.M so special care is required during these Peak Hours for Network maintenance. Also the call rate should be adjusted during these Peak Hours of Traffic. Similarly, the Non Peak Hours for Network Traffic are 11 A.M to 12 P.M. Marketing campaigns should address these Hours and different rates are offered for these Off Peak hours for better utilization of network and hence to produce more revenue for business.

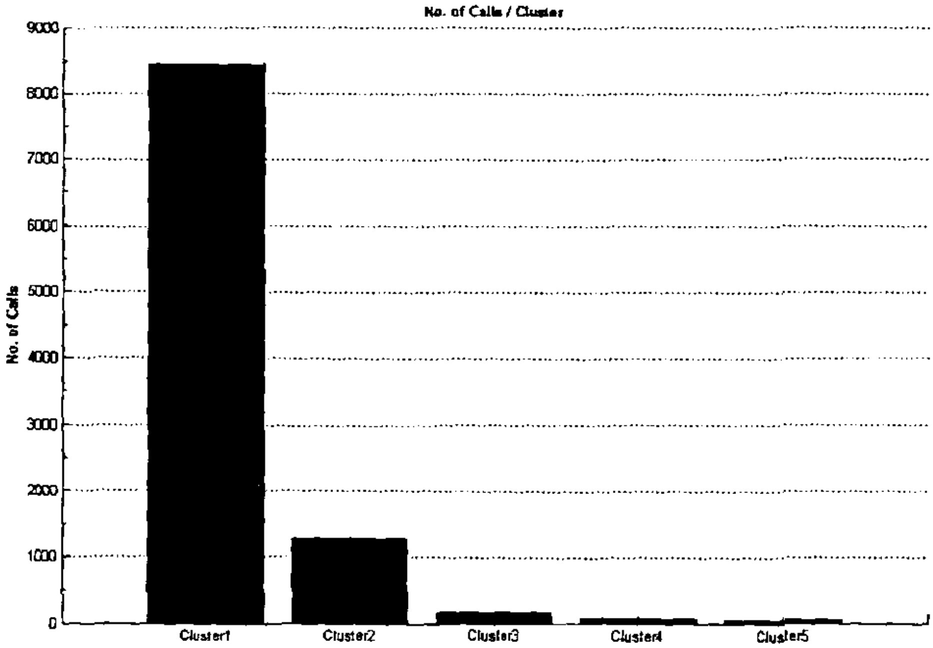

Figure 19: ART based Clusters for Time of Call

Figure 19 shows the categorization formed by ART algorithm. Although the clustering / segmentation done by this algorithm is totally different from other two but the analysis of results shows the same. ART categorizes the input data into 5 clusters of Subscribers with respect to their Call Timings based on their 15 days CDR data. It shows that the 10,000 subscribers are segmented as follows:

Cluster	No. of Subscribers	Cluster Explanation
1	8,100	Most Busy Network w.r.t Traffic
2	1,300	Average Network w.r.t Traffic
3	300	Not Busy of Network w.r.t Traffic
4	200	Not Busy Network w.r.t Traffic
5	100	Not Busy Network w.r.t Traffic

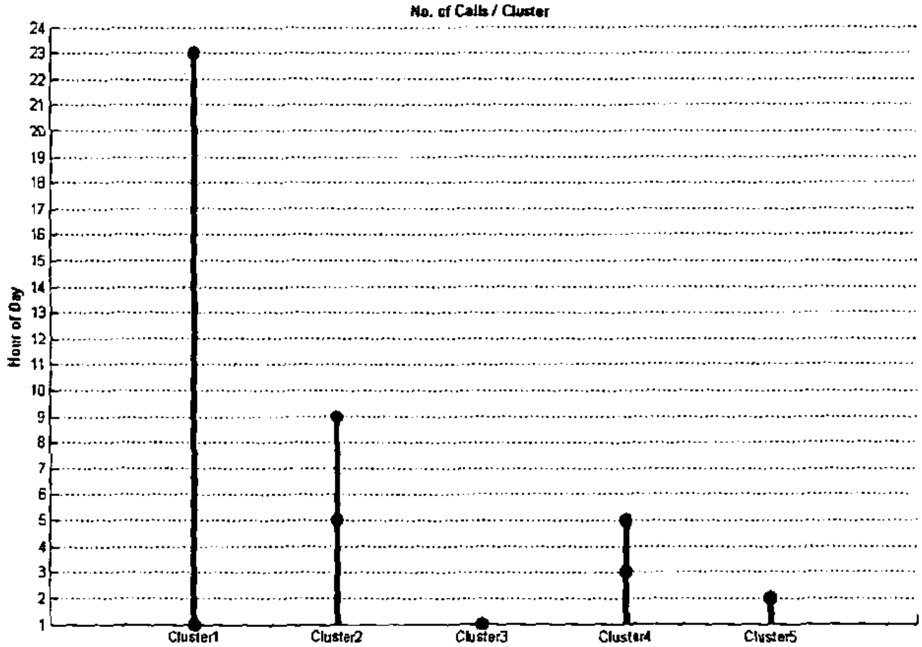

Figure 20: ART based Cluster Ranges for Time of Call

The range of Hours w.r.t Call Time in different clusters by ART is shown in Figure 20.

Cluster	**Range of Hours**	**Cluster Explanation**
1	12 A.M. to 11 P.M	Most Busy Network w.r.t Traffic
2	5 A.M. to 9 A.M	Average Network w.r.t Traffic
3	12 A.M. to 1 A.M	Not Busy Network w.r.t Traffic
4	3 A.M. to 5 A.M	Not Busy Network w.r.t Traffic
5	12 A.M. to 1 A.M	Not Busy Network w.r.t Traffic

The analysis obtained by ART algorithm is not satisfied with given data because of input measures selected for analysis with this algorithm.

5.3.3 Trend Analysis

The following are the results for trend analysis.

5.3.3.1 Trend Analysis of Revenue Data

Trend Analysis w.r.t Service Package and Subscribers

Figure 21 shows trend of most revenue generating subscribers that belong to Cluster No. 1 (Figure 9). Further analysis of these customers shows that most of these subscribers contain service Package No. 7, 5, 2 & 10. So we can say that most revenue generating customers belong to these service packages. There is a potential for marketing campaign for these packages.

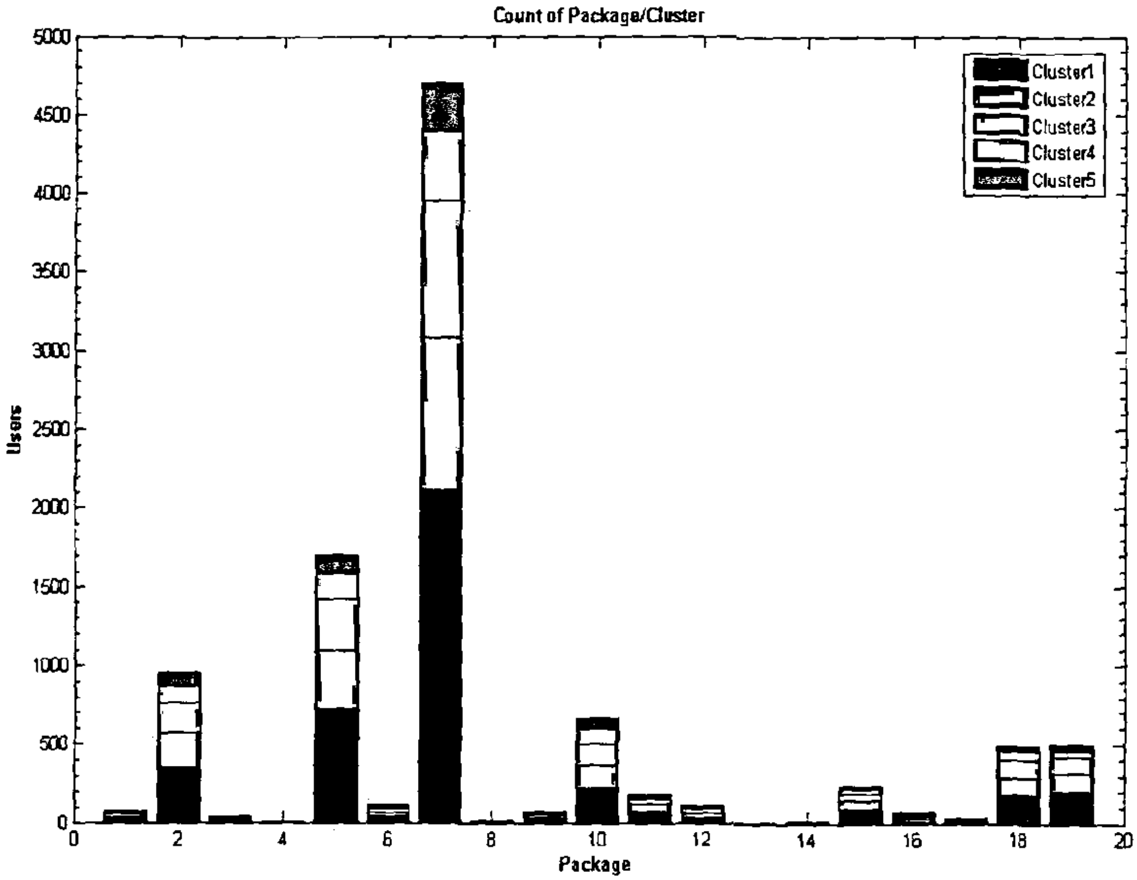

Figure 21: SOM based Trends of Most Revenue generating Packages

Figure 22 shows trend of most revenue generating subscribers that belong to Cluster No. 5 (Figure 11). Further analysis of these customers shows that most of these subscribers contain service Package No. 8, 6, 3, 11, 20 & 21. So we can say that most revenue generating customers belong to these service packages. There is a potential for marketing campaign for these packages.

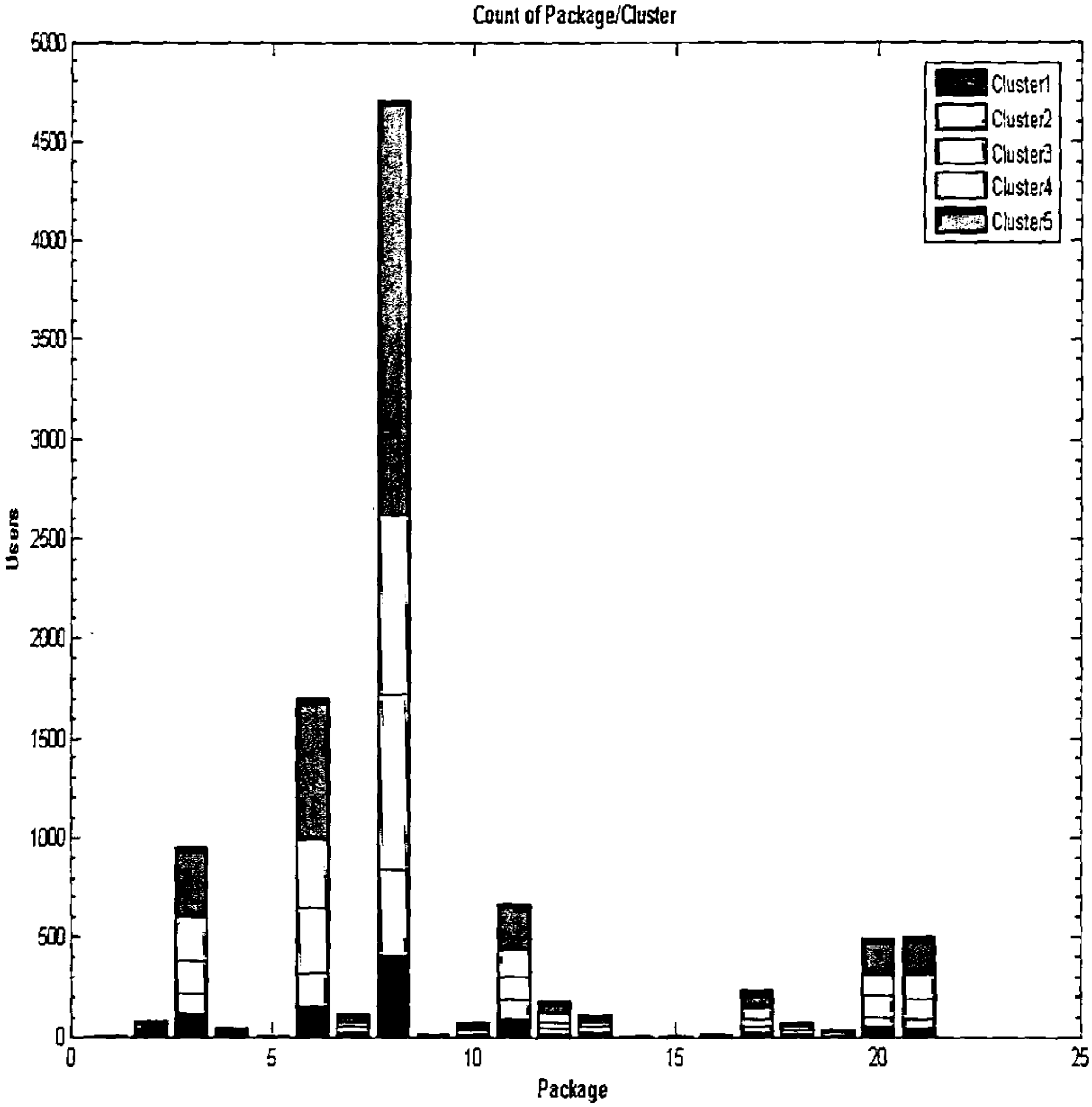

Figure 22: LVQ based Trends of Most Revenue generating Packages

Figure 23 shows trend of most revenue generating subscribers that belong to Cluster No. 1 (Figure 13). Further analysis of these customers shows that most of these subscribers contain service Package No. 10, 7, 3, 14, 32 & 33. So we can say that most revenue generating customers belong to these service packages. There is a potential for marketing campaign for these packages.

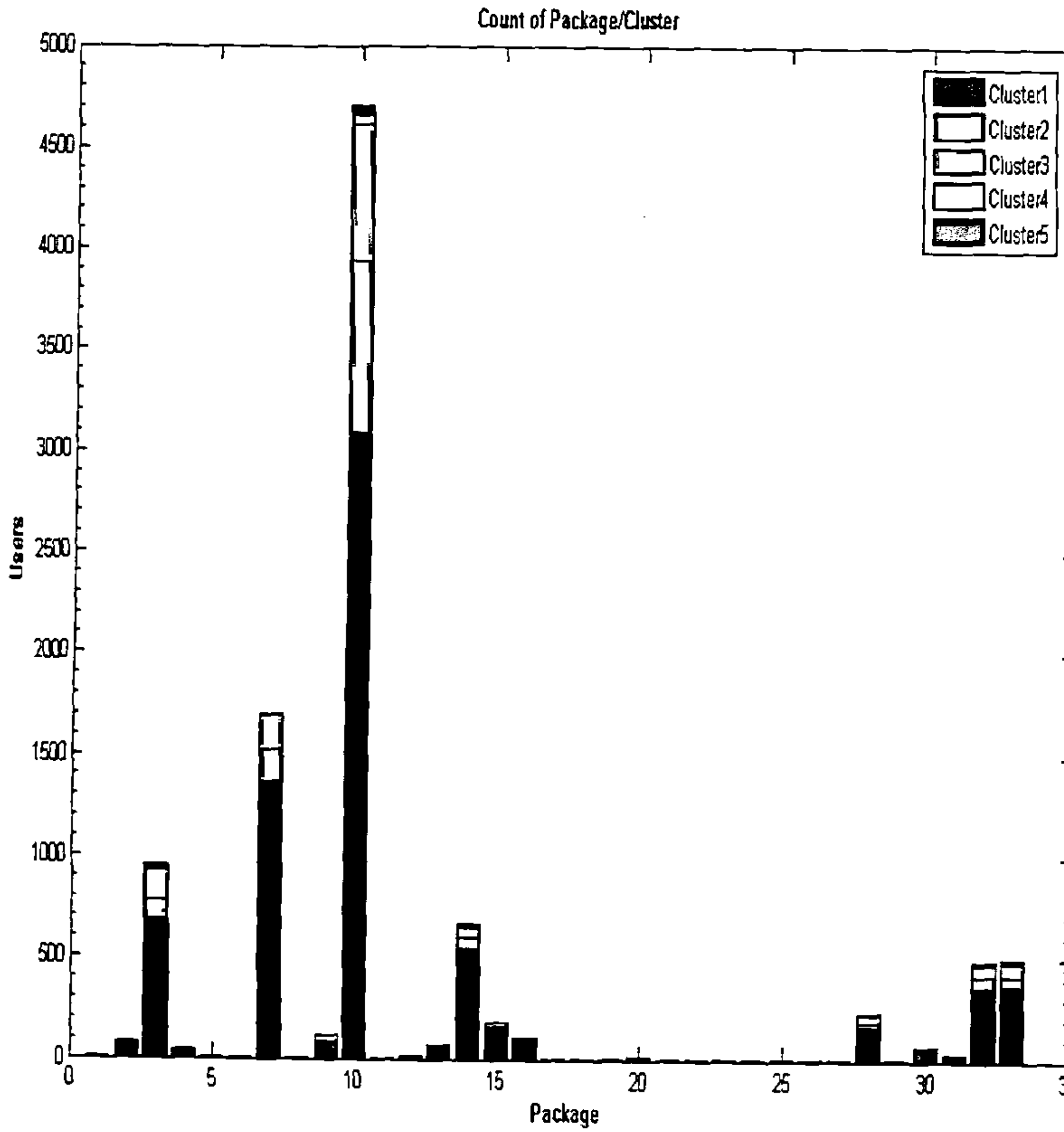

Figure 23: ART based Trends of Most Revenue generating Packages

Trend Analysis w.r.t City and Subscribers

Figure 24 shows trend using SOM algorithm of most revenue generating subscribers that belong to Cluster No. 1 (Figure 9). Further analysis of these customers shows that most of these subscribers belong to City No. 41, 52, 63 and 94. So we can say that most revenue generating customers belong to these cities. There is a potential for marketing campaign for these cities.

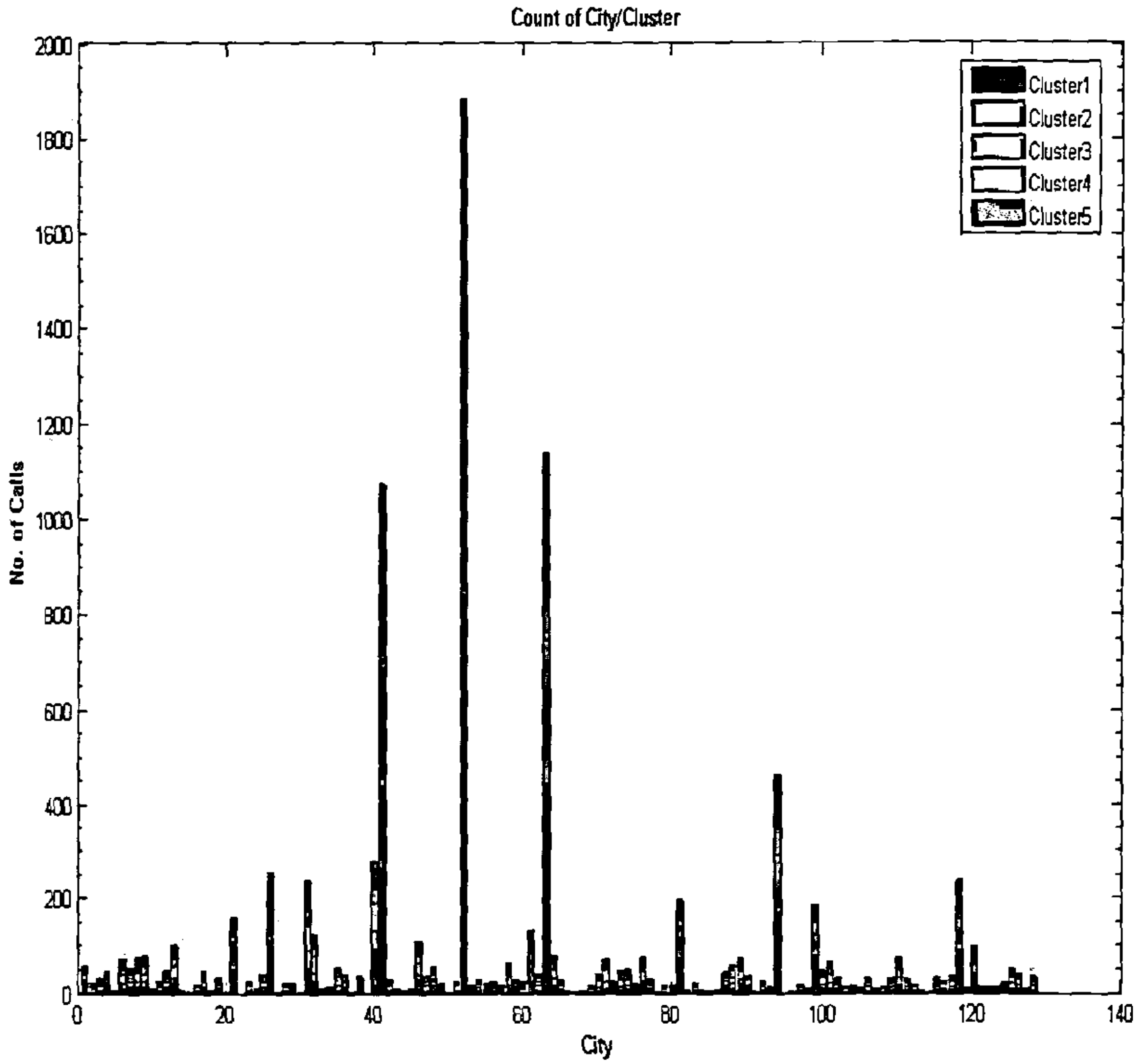

Figure 24: SOM based Trends of Most Revenue generating Cities

Figure 26 shows trend using ART algorithm of most revenue generating subscribers that belong to Cluster No. 1 (Figure 13). Further analysis of these customers shows that most of these subscribers belong to City No. 43, 54, 65 and 99. So we can say that most revenue generating customers belong to these cities. There is a potential for marketing campaign for these cities.

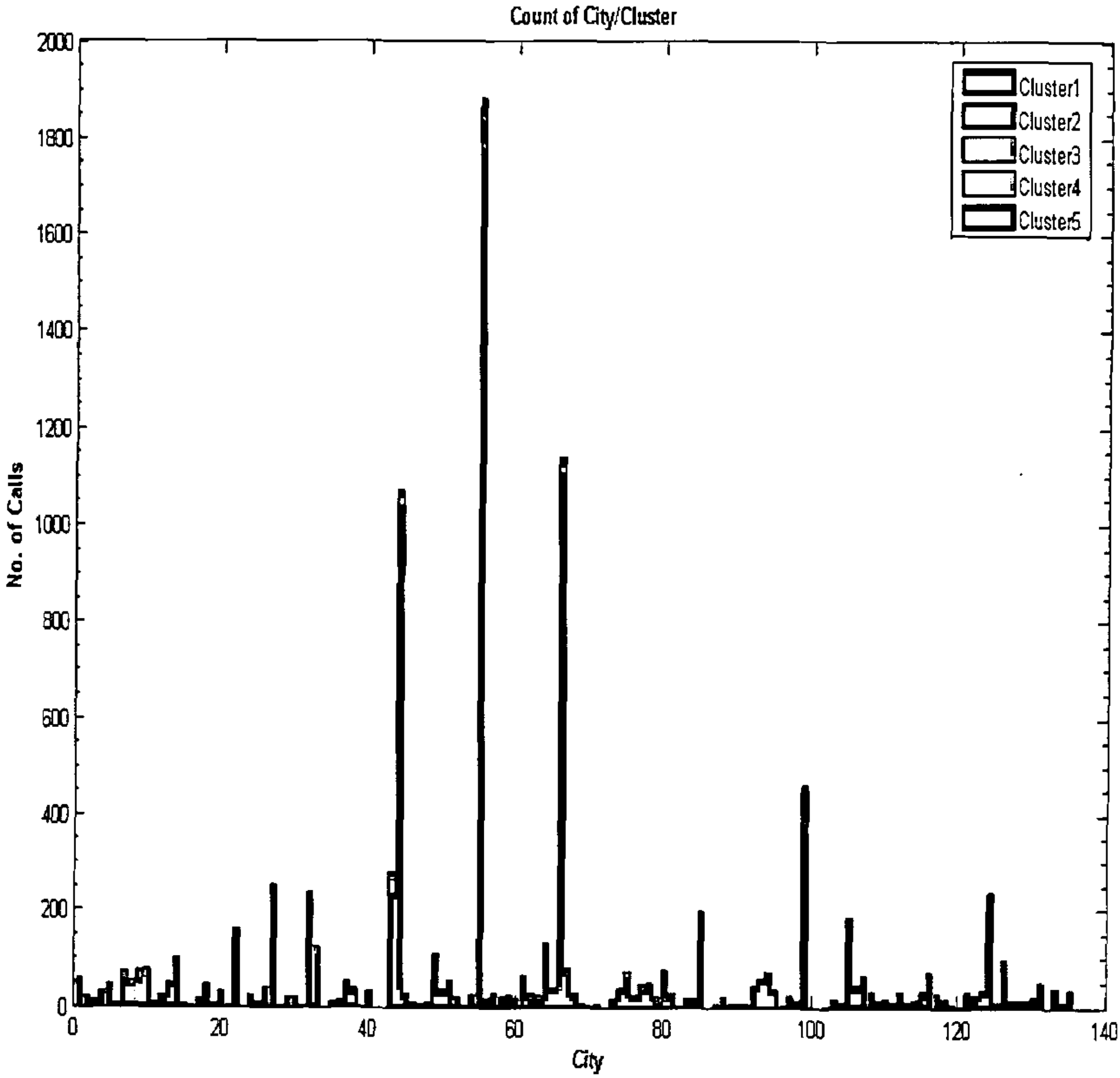

Figure 26: ART based Trends of Most Revenue generating Cities

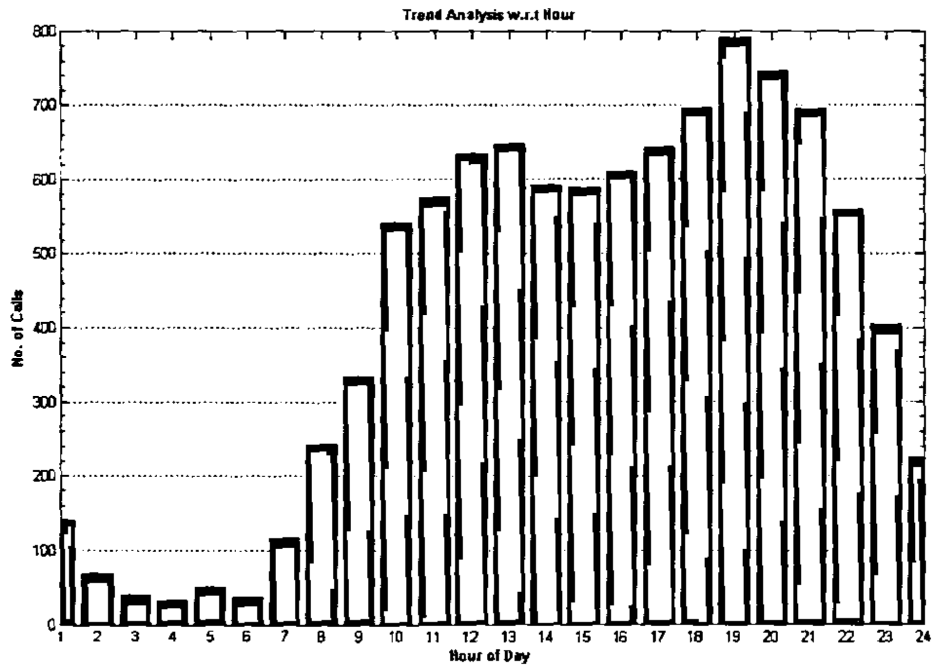

Figure 28: LVQ based Trends of Most Busy Hours

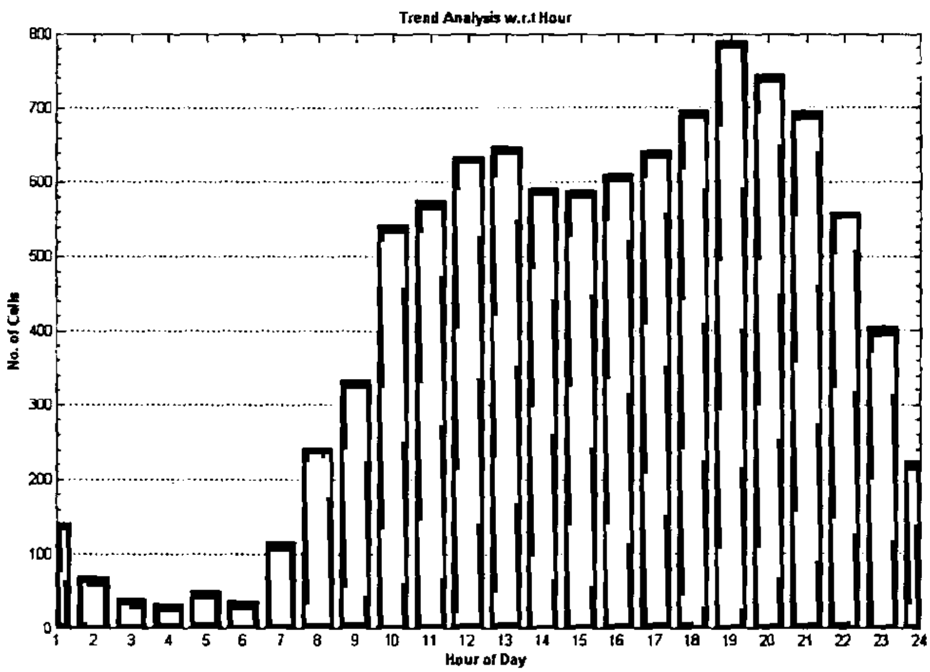

Figure 29: ART based Trends of Most Busy Hours

Trend Analysis w.r.t City and Total Call Time

Figure 30 shows trend using SOM algorithm of most busy hours of Telecom Network Traffic w.r.t City. Further analysis of Network Traffic shows that there are specific cities that are generating much traffic on network on Peak Times (06 P.M to 09 P.M) (Figure 27). These cities are City No. 51, 64 and 41. Similarly there are specific cities that are not contributing as much traffic in Peak Hours. On the basis of this trend, we say that there is potential for marketing campaigns for the utilization of this time period in specific cities

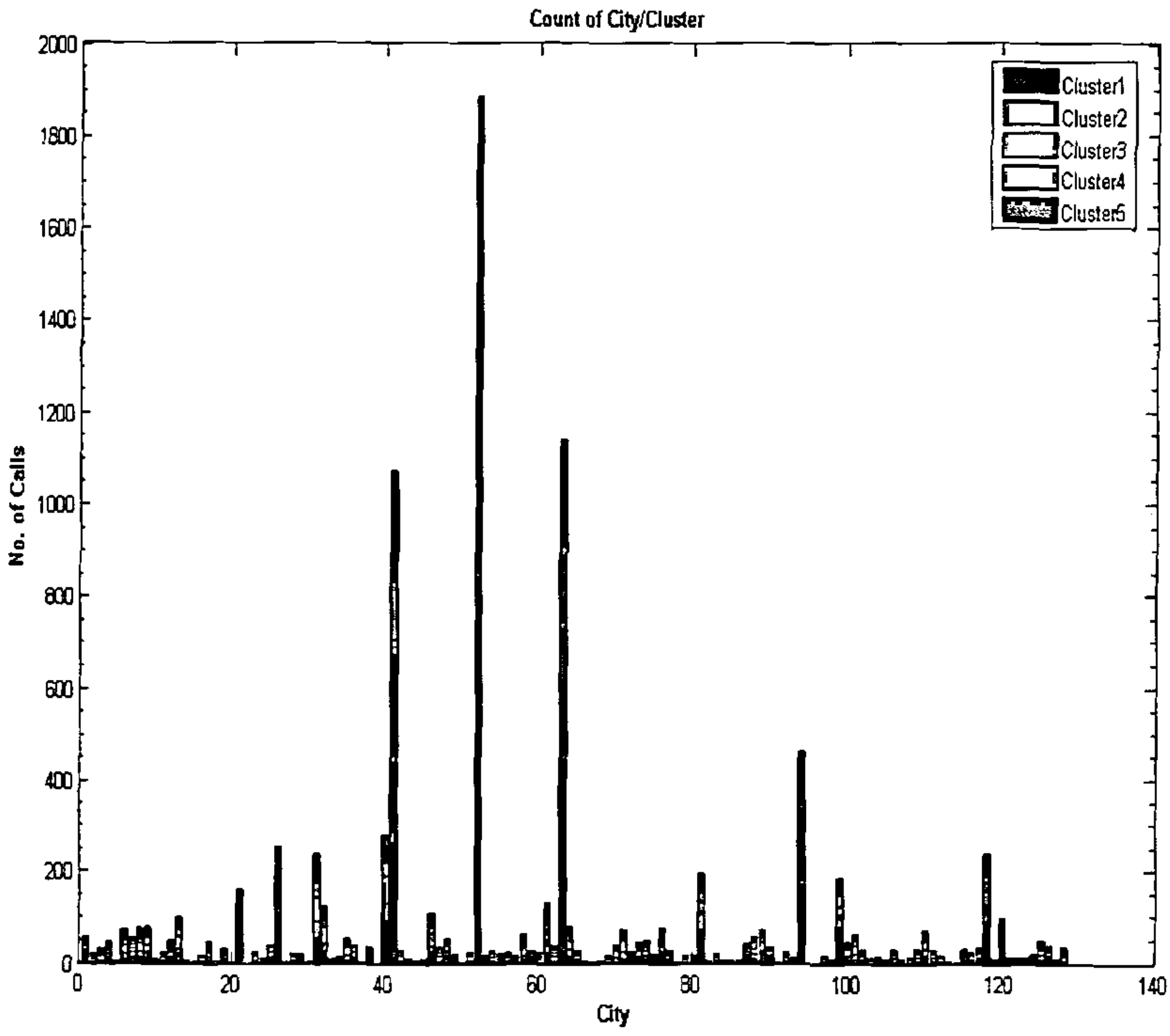

Figure 30: SOM based Trends of Cities with Most Busy Hours

Figure 31 shows trend using LVQ algorithm of most busy hours of Telecom Network Traffic w.r.t City. Further analysis of Network Traffic shows that there are specific cities that are generating much traffic on network on Peak Times (06 P.M to 09 P.M) (Figure 28). These cities are City No. 51, 64 and 41. Similarly there are specific cities that are not contributing as much traffic in Peak Hours. On the basis of this trend, we say that there is potential for marketing campaigns for the utilization of this time period in specific cities

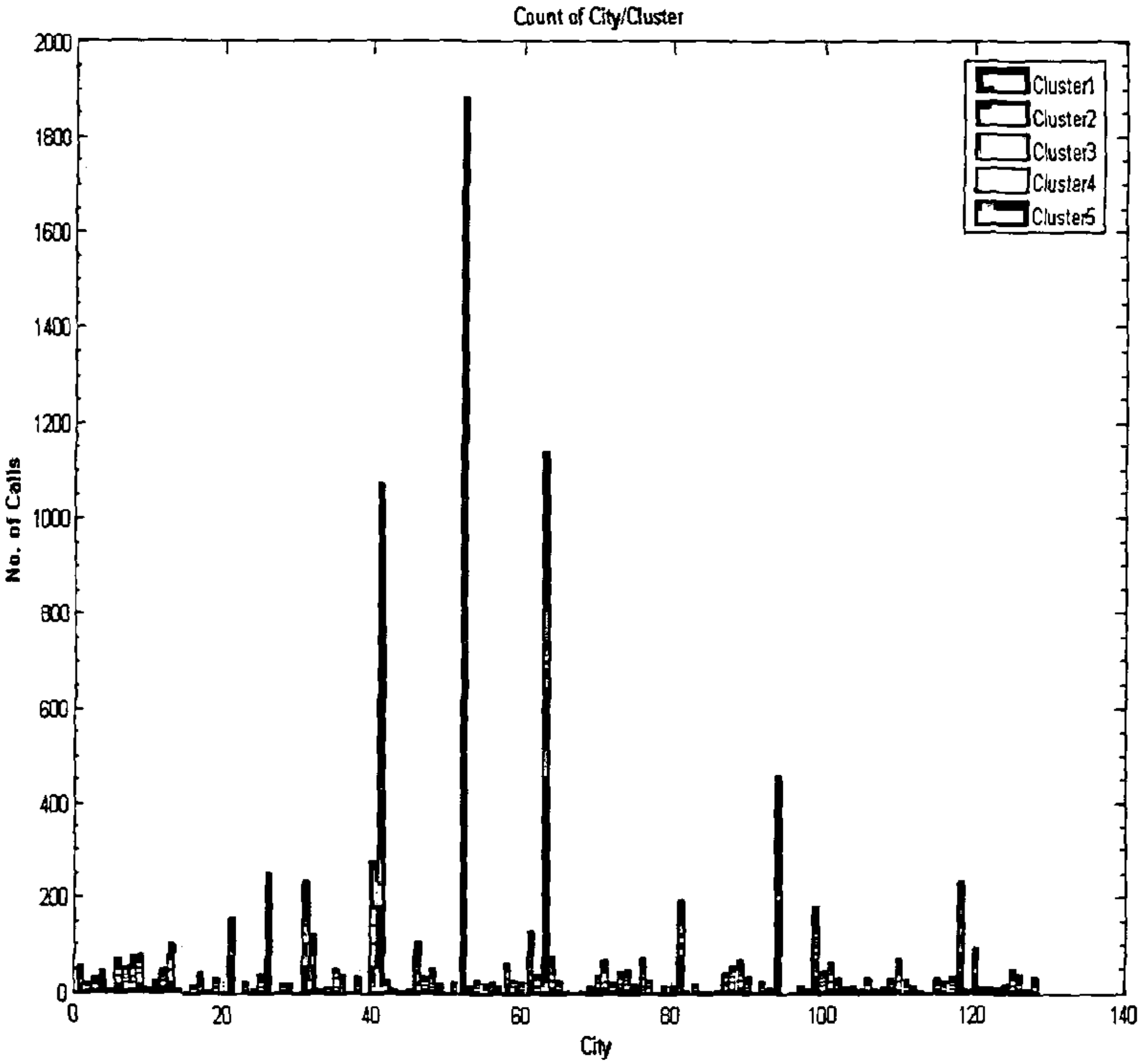

Figure 31: LVQ based Trends of Cities with Most Busy Hours

Figure 32 shows trend using ART algorithm of most busy hours of Telecom Network Traffic w.r.t City. Further analysis of Network Traffic shows that there are specific cities that are generating much traffic on network on Peak Times (06 P.M to 09 P.M) (Figure 29). These cities are City No. 55, 66 and 44. Similarly there are specific cities that are not contributing as much traffic in Peak Hours. On the basis of this trend, we say that there is potential for marketing campaigns for the utilization of this time period in specific cities

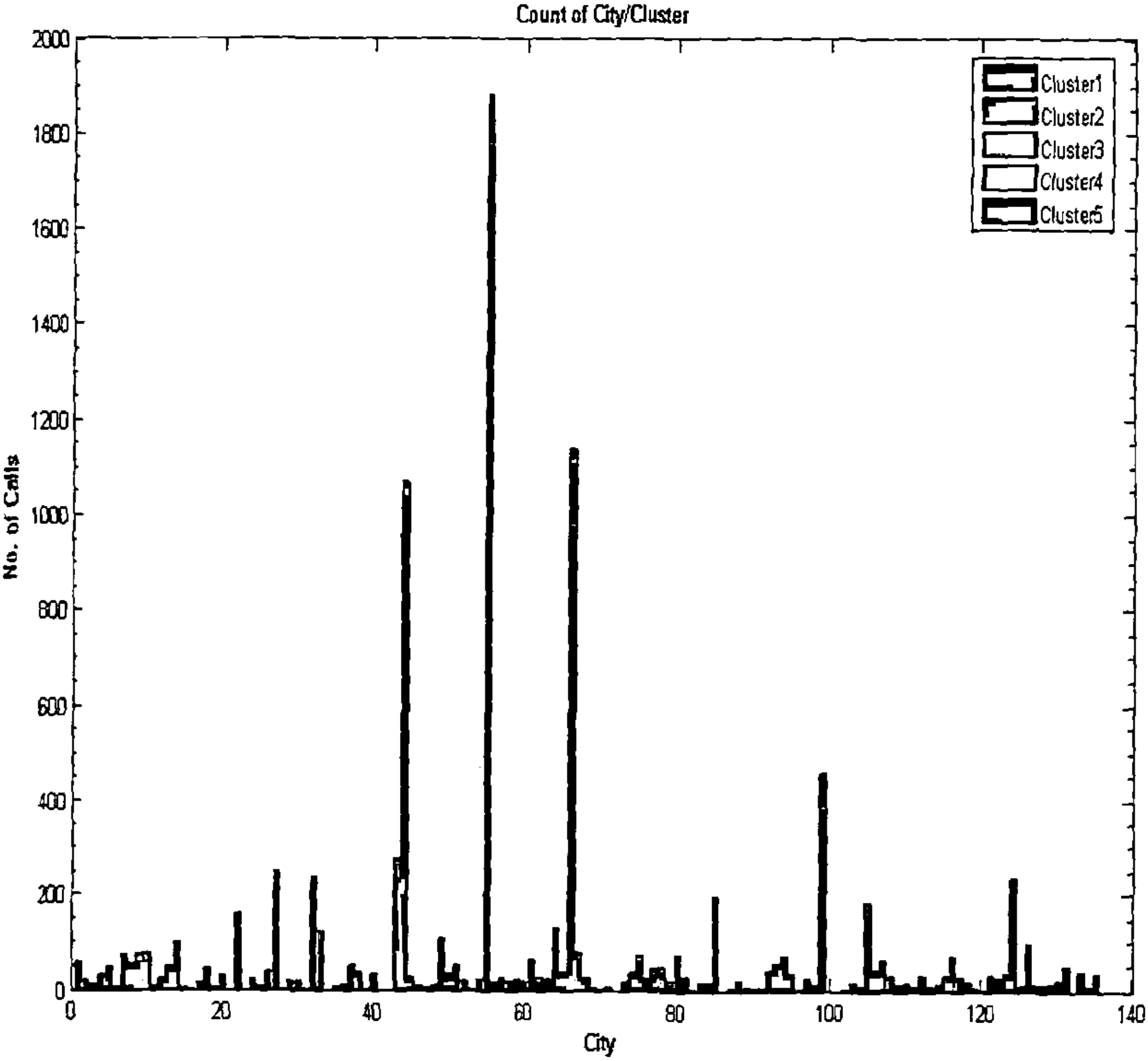

Figure 32: ART based Trends of Cities with Most Busy Hours

Chapter 6

CONCLUSION & OUTLOOK

6. Conclusion and Outlook

6.1 Conclusion

The primary objective of this project has been to investigate the unsupervised learning potentials of novel neural networks for the trend analysis of calls made by users over a period of time in a mobile telecommunication network. And, build a tool for the visualization of natural groupings in the multidimensional data and trends from these groupings. A requirement was to base this tool on unsupervised neural network algorithms, where similar items from the data space appear close to each other, thus inherently clustering the data.

Specifically, this study provides a comparative analyses and application of Self Organizing Maps (SOM), Learning Vector Quantization (LVQ) and Adaptive Resonance Theory (ART) competitive neural networks algorithms to user call data records in order to conduct a descriptive data mining on users call patterns.

Our investigation shows the learning ability of both neural networks to discriminate user call patterns with the SOM neural network algorithm providing a better discrimination of call patterns than the ART algorithm.

Following are the tables showing the comparison for the three different algorithms.

Comparison of Parameter values for three types of Unsupervised Neural Network algorithms

Algorithm	No. of Input Records	Parameter			Accuracy
		Learning Rate (α)	Total Iterations	No. of Clusters	
SOM	100	0.6	50	5	80%
SOM	1000	0.6	75	5	90%
SOM	10000	0.6	100	5	100%
SOM	20000	0.6	> 100	5	100%

Algorithm	No. of Records		Parameter			Accuracy
	Training	Input	Learning Rate (α)	Total Iterations	No. of Clusters	
LVQ	50	1000	0.1	50	5	80%
LVQ	100	5000	0.1	75	5	90%
LVQ	500	10000	0.1	100	5	100%
LVQ	1000	20000	0.1	> 100	5	100%

Algorithm	No. of Input Records	Parameter				Accuracy
		Learning Rate (α)	Noise (θ)	Vigilance (ρ)	No. of Clusters	
ART	100	0.39	0.1	0.9	3	80%
ART	1000	0.39	0.5	0.5	4	80%
ART	10000	0.39	0.1	0.9	4	90%
ART	20000	0.39	0.1	0.9	5	100%

The comparative analysis findings are as follow:

SOM is much better in term of results. The reason is that SOM categorizes data basically with respect to first component of input vector.

The degree of correctness of categorization with LVQ depends on the number and type of learning data. Also result show that LVQ is faster in term of execution time. The reason behind this is that the out put of SOM is given as input to LVQ which increases its efficiency.

ART is good in term of results. ART algorithm depends on all the components of input vector. In ART, parameter θ have a significant effect on the performance of the net. The value of vigilance parameter ρ determines how many clusters will be formed. It allows increase in number of clusters only if required

Following table shows the results of the three algorithms showing their efficiencies and accuracy of clusters.

Results showing Execution Time and %age of Accuracy of Clusters of three types of Unsupervised Neural Network algorithms

Algorithm	No. of Input Records	Execution Time (Approx.)	Accuracy of Clusters
SOM	20000	25 Minutes	100%
LVQ	20000	15 Minutes	97%
ART	20000	20 Minutes	90%

This table shows that although total number of input records for all the three algorithms were same i-e 20,000 but time taken by them to make clusters is different. SOM took maximum time but its result is 100%. LVQ took less time for execution as compared to SOM because it gets its input from SOM (1000 records) and therefore its training gets completed in very short time. Remaining time which it takes for execution is to make clustering on the basis of this training. Also ART took less time than SOM but its clustering is not as good as of SOM.

Chapter 7

REFERENCES

7. References

[Abidogun (2005)] O. Abidogun, "Data Mining, Fraud Detection and Mobile Telecommunication: Call pattern Analysis with Unsupervised Neural Networks," M.Sc. thesis, University of the Western Cape, Bellville, Cape Town, South Africa, 2005

[Carpenter & Grossberg (1987a)] Carpenter.G.A., & Grossberg.S. (1987a). A massively parallel architecture for a self-organizing neural pattern recognition machine. Computer Vision, Graphics, and Image Processing. 37. 54-115.

[Carpenter & Grossberg (1987b)] Carpenter.G.A., & Grossberg.S. (1987b). ART 2: Self-organization of stable category recognition codes for analog input patterns. Applied Optics. 26. 4919-4930.

[Carpenter et al. (1991)]. Gail A. Carpenter, Stephen Grossberg, and D.B. Rosen. Art2-a: An adaptive resonance algorithm for rapid category learning and recognition. *Neural Networks*, 4:493-504, 1991

[Cerny (2001)] P. Cerny. "*Data mining and Neural Networks from a Commercial Perspective*", ORSNZ Conference Twenty Naught One, 2001.

[Clementine (2008)] "Clementine--A Data Mining Toolkit" [accessed 23.01.2008]<http://www.mip.com.au/clemkit.html>

[Craven and Shavlik (1998)] M. Craven and J. Shavlik. Using neural networks for data mining, Future Generation Computer Systems, vol. 13 (Special issue on Data mining), pp. 211--229, 1998.)

[CRISP-DM (2007)] CRISP-DM project [accessed 18.09.2007] <http://www.crisp-dm.org/Process/index.htm >

[Dimitrios and Christos (1997)] Siganos, Dimitrios & Stergiou, Christos, Neural Networks, 12.5.1997 [accessed 12.03.2008] < http://www-dse.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html>

[Dolnicar and Leisch. (2001)]	Dolnicar, S., & Leisch, F. (2001). Behavioral market segmentation of binary guest survey data with bagged clustering. In G. Dorffner, H. Bischof, & K. Hornik (Eds.), ICANN 2001 (pp. 111–118). Berlin: Springer-Verlag.
[Fausett (1994)]	Laurene Fausett, Fundamentals of neural networks: architectures, algorithms, and applications, Prentice-Hall, Inc., Upper Saddle River, NJ, 1994
[Frawley et al. (1991)]	W. Frawley and G. Piatetsky-Shapiro and C. Matheus, 1991: "Knowledge Discovery in Databases: An Overview of *Knowledge Discovery In Databases*," AAAI Press/MIT Press, Cambridge, MA. pp. 1-30.
[Gary (2005)]	Gary M. Weiss, 2005: Data Mining in Telecommunications The Data Mining and Knowledge Discovery Handbook 2005: 1189-1201
[Ghahramani (2004)]	Z. Ghahramani. Unsupervised learning. In O. Bousquet, G. Raetsch, and U. von Luxburg, editors, Advanced Lectures on Machine Learning, volume LNAI 3176. Springer-Verlag, 2005.
[Grossberg (1976a)]	Grossberg. S. (1976a). Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors. Biological Cybernetics. 23. 121-134.
[Grossberg (1976b)]	Grossberg. S. (1976b). Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions. Biological Cybernetics. 23. 187-202.
[Hillary et al. (2001)]	Hilary K. Browne , William A. Arbaugh , John McHugh , William L. Fithen, A Trend Analysis of Exploitations, Proceedings of the 2001 IEEE Symposium on Security and Privacy, p.214, May 14-16, 2001

[Zhang et al. 2004] D. Zhang, S. Chen, and Z.-H. Zhou. Fuzzy-kernel Learning Vector Quantization. Proc. 1st Int. Symp. on Neural Networks (ISNN'04, Dalian, China), LNCS 3173, 180–185. Springer-Verlag, Berlin, Germany 2004