MINING CALL CENTER PERFORMANCE DATA

By

RUIYUAN GUO

Master of Education

Beijing Normal University

Beijing, China

July 1991

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements of the Degree of
MASTER OF SCIENCE
August 2003

MINING CALL CENTER PERFORMANCE DATA

Thesis Approved:

Thesis Advisor

Dean of the Graduate College

ACKNOWLEDGMENTS

I would like to express my grateful appreciation to my advisor, Dr. Marcin Paprzycki for his advice, encouragement, patience and kindness throughout my thesis study. I would also like to express deep appreciation to Dr. Ajith Abraham for his warm, constant and valuable guidance. I would also extent sincere thanks to Dr. Dursun Delen for his suggestions and comments.

I would also like to express my deep thanks to all my family members, especially to my parents and my husband, for their love and support. My thanks also go to all my friends, especially to Miaomei Chong, for their constant help and encouragement.

TABLE OF CONTENT

LIST OF TABLE

Table

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Call center performance depends on the performance of its customer service representatives (CSRs) and call traffic regulations. The performance data of call centers include quality assessment data, time management of the calls, and business processing aspects during the calls. The team leaders of call centers manage call centers, deliver training courses, and reward CSRs to improve the performance of CSRs according to their performance results.

The previous data mining research related to call centers was focused on the data related to customers. The research analyzed the behaviors of customers and its aim was customer satisfaction improvement. However, there is not much research found on mining the data of performance of call center representatives.

The aim of our research is to apply data mining techniques to the performance results collected from five call centers of a large insurance company in an attempt of discovering interesting information that can help improve the performance of individual representatives and thus the call centers.

The remaining parts of this thesis are organized as follows: Chapter II reviews the literatures related data mining related to call centers. Chapter III states the objectives of the research. Chapter IV introduces the design and implementation of the proposed model. Chapter V analyzes the results of deployment. Chapter VI aims the conclusion of the study.

CHAPTER II

REVIEW OF LITERATURE

In the past decade, the call center industry has been one of the most rapidly growing industries. "The Call Center Association defines call centers as a physical or virtual operation within an organization in which a managed group of people spend most of their time doing business by telephone, usually working in a computer – automated environment" [15]. The service performance of call centers directly impacts customer retention and loyalty, thus impacting the revenue generation. Call center managers have endeavored to improve the performance of call centers by using new technologies that became available in recent years. Thus, customer relationship management has become a solution.

Customer Relationship Management (CRM) has been the focus of a substantial body of research. In fact, CRM is not really new. When new technologies, such as the Internet, monitoring systems and data mining, become available, CRM adopted them and reached new level of effectiveness [32]. Technologically, modern CRM is based on the use of data mining to identify customer preferences and behavior, as well as customer service representatives (CRS) behavior patterns. Identifying the behavior patterns and finding the facts that impact the service quality can help improving the performance of service and further improving the customer satisfaction. For example, the possible facts that could affect the performance of a representative are training received, coaching, and the complexity of products. The possible fact can even be the business culture and model of the whole call center.

However, current research attempting to mine call center data is focused mostly on customer data, such as age, profession, salary, call information, etc. In the summary of past research we concentrate on analyzing the ways that data was collected and the technologies used in the previous research.

2.1 Data Collection

The customer data used for call center related mining is mainly from three sources: customer databases, web usage, monitoring system.

2.1.1 Customer databases

Customer databases store basic data related to customers such as customer age, address, profession, amount and products purchased, and information about past interactions with the call centers such as calling date, calling reason, etc. If customers send in requests by emails instead of calls, the emails are stored into the database as well [7].

2.1.2 Monitoring System

Monitoring systems record calls and also capture the screen of CSR's PC. A group of people in call centers evaluates or translates customer's behavior and puts the evaluation results into a database for mining customer data.

There are two different ways to report calls by the monitoring systems. Some monitoring systems start to record when CSRs pick up customer's calls. Some monitoring systems record once customer's call is connected by interactive voice response (IVR). The second way could provide the data of how many times a customer is transferred and how long a customer has been waiting.

Tan et al. collected all requests from customers and put into a call tracking database. The database contained both fixed-format fields and free-text fields. The fixed-format fields store the information such as case number, type of problem encountered, product ID, and the amount of time spent on the request, etc. The free-text field contained a case description written by the engineers who handled the case [36].

2.1.3 Web Usage

Web sites are published by call centers for customers to obtain information, purchase products or pay for the services. Web usage contains valuable data about customer behavior.

Web usage data can be collected at server side, client side, or at the proxy servers. Most researchers use server log files. However, the server-side data is considered not reliable and inefficient. Cache hits can be missed. Information passed through POST method is typically not available. Client-side collection or web proxy server-side collection can be used as makeup in the situation when cache hit missed.

Client-side collection has advantages over server-side collection that it allows to eliminate some of the above described unreliability and inefficiencies. Client-side collection can be implemented by a modified browser or a client agent. However, it can post a lot overhead in addition to privacy concerns and requires user's cooperation [10].

A web proxy is an intermediate level between client and server. Proxy traces can provide actual HTTP requests [10].

2.2 Mining Technologies and Results

According to the methods of data collection and the nature of the data, mining technologies are slightly different.

2.2.1 Mining Data Collected from Monitoring System and Customer Database

Some venders of monitoring system such as eTalk and GartnerGroup built data mining tools into their monitoring systems. However, users of monitoring system are intended to be non-data-mining experts, such as supervisors and managers. A call center supervisor mines the pattern of customers by asking "what if" questions [13]. Dilauro et al. found that customer could get frustrated if the waiting time is longer than the talking time. Also they found that the more a call was transferred, the more frustrated the customer could get.

2.2.1.1 Predictive modeling

Predictive modeling such as decision-tree and neural network can be used to predict customer behavior. Quaero LLC used CHAID, CART analyses along with multivariate regression of neural network and classified customers into four clusters: high current value with high potential value, low current value wither high potential value, high current value with low potential value, and low current value with low potential value. A set of available customer characteristics are mapped as input to these four clusters. [19]

2.2.1.2 Textual data mining

Textual data mining of customer mining is another mining technology applied call centers. Busemann et al. classifies e-mail request from customers based on shallow text processing and machine learning techniques. SMES, an information extraction core system by Busemann et al., carried out shallow text processing, SMES had three components: a text tokenizer, a lexical processor and a chunk parser. The tools Busemann et al. used for the mining were Lazy Learning of MLC++ library by Kohavi and Sommerfield, Symbolic Eager Learning, Support Vector Machines which is SVM_Light by Joachims, Neural Networks of Learning Vector Quantization by Kohonen. The system repeatedly chose an e-mail request and found a solution. An agent adjusted the solution if it was wrong. They found that the accuracy of the system was 73% [7].

In their research, Tan et al. removed irrelevant sentences with free-text field, changed different morphological words into same lexical unit, removed spelling mistakes, etc. Then tokenizer was used to break the free-text sentences into smaller tokens. The mining was done on tokens by using the C4.5 decision trees and the Naïve Byesian. They found that their model could help to classify the duration of a service case and also better classify the service calls [36].

Finally, some researchers attempted at audio data mining. ScanSoft is the leading product of audio mining area [25]. ScanSofter uses context-free-grammar to parse the speech. A method called Sequence Package Analysis creates caption text. The mining data is actually the caption text. Neustein found that this audio data mining strategy could predict some of the subtle key words in callers' speech, such as early warning signs of caller frustration [25].

2.2.2 Web Usage Mining

Web usage mining has been focused on three areas, association, clustering, and path analysis. The research described below is not necessarily call-center related. However, the web usage mining of call-center related data is essentially the same as other web usage mining: companies present their products on the web; customers navigate the web, view the information and buy the products. The review below will be focusing on the techniques the researches used that could help the call center related web using mining.

Association rules are used to find related pages that are referenced together in a single session.

Tan et al. applies a new data mining technique called indirect association to sequential data of web usage. A pair of attributes A and B is said to be indirectly associated via a mediator set M. They first generate all large itemsets by using standard frequent itemset generation algorithm. Then they associate each itemset to the candidate pairs of indirect association. The results indicate that the new technique is capable of distinguishing the interest of web users who are traversing a web site [35].

Clustering is used to group users or pages that have similar characteristics. There is research that applies fuzzy logic to cluster web usage data.

To cluster sessions, Joshi etc. developed a "Fuzzy c-Medoids Algorithm" based on Fu's heuristic algorithm and Least Trimmed Squares idea. The time complexity of the algorithm is O(kn) where k is a low integer. The research result shows that the fuzzy clustering process could create better session profiles [18].

Path discovery and analysis is another method used in web usage. The research of Chen etc formats data into a set of maximal forward references. They found large reference sequences that represent frequent reference sequences among maximal forward references by using both full scan algorithm and selective scan algorithm. They found that Selective scan has good performance on CPU and memory usage. However, the researcher did not mention comparison result of the effectiveness.

CHAPTER III

RESEARCH STATEMENT

A relationship is built between two objects or two ends. One object can never construct a relationship. So is true in customer relationship management of call centers. Customer relationship management of call center is managing relationship between customer and customer service representatives. A call center is never able to achieve its CRM objectives and improve the customer satisfaction without ensuring the representative satisfaction first. The above research summarized was concentrated on customer data and did not consider mining performance data of CSR. The objective of the research is to apply data mining in CSR performance results of call centers of an insurance company and find interesting characteristics that can help understanding and improving the performance of CSRs.

3.1 Business Objectives

Call centers use the performance results to reward CSRs monthly, quarterly and yearly. However, due to various reasons in certain time period, monitoring cannot be done on certain CSRs. The CSRs who are missing quality score are given a "met" score to compete with other CSRs. Call center management team is seeking a better way to handle this kind situation.

The main objective of the research is to build a model that can predict monthly quality scores of CSRs whose quality scores are missing in a given month.

The performance of a call center directly affects customer satisfaction and retention. The performance of a call center can be measured in two dimensions: efficiency, which is also, called productivity, and effectiveness, which is also, called quality [32]. Call center management team puts a lot effort to improve customer service quality while balancing the call volume. The second business objective of the research is, to find the relationship between quality and productivity and to find the improvement opportunities of CSR performance for training and coaching, further improve customer satisfaction and retention.

3.2 Data Mining Goals

The research uses the representative's actual performance data of the call centers of an insurance company. The mining goal is to build models and predict quality. The research will compare the performance of different models by applying Multi-Layer Perceptrons, CART, SVM-Light, ANN-CART hybrid and Linear Model to the same dataset from call center performance database. The research is also to analysis the relationship among the attributes to find the improvement opportunities for training.

CHAPTER IV

DESIGN AND MODELING

This chapter will introduce the dataset used in this study and summarize the procedures applied to prepare the data. The chapter will also present the procedures of model building.

4.1 Data Understanding

The data used in the study is one year's actual data from the performance evaluation database of the call centers of an insurance company. Each customer service representative has a record for each month. The attributes of each record consist of two categories: monitoring results and time management.

4.1.1 Monitoring Results Collection

Each CSR answers ten to sixty calls everyday. The phone calls are randomly recorded when CSRs customer calls and the monitoring system constantly keeps up to ten calls for each CSR. A group of evaluators review the recorded calls and evaluate the calls according to certain business criteria. The results of the evaluations are the monitoring results.

In the insurance company from which the data was obtained, there are two main attributes of monitoring results: customer service satisfaction and business need satisfaction.

Customer service satisfaction score is an aggregate result of evaluation based on eleven questions for all products handled by the call centers. Typical questions of customer service satisfaction are "did a CSR thanked the customer for calling the company" or "did a CSR asked what else they can help customer". Evaluators evaluate each question by assigning an integer number from 0 to 5. 0 represents that the question is not applicable to the call. 1 indicates that the CSR did not meet the expectation. A 2 indicates that the CSR met to some degree, which is also said to be "met some". A 3 indicates that the CSR met the expectation. A 4 is that the CSR exceeded the expectation. A 5 represents that the CSR far exceeded the expectation. These results are then aggregated to a score representing the total level of meeting customer service satisfaction criterion.

For example, a call is recorded in the system. An evaluator reviewed the call and found that only three questions out of eleven are applicable to the call. The evaluator marked three questions 3, 4 and 1 according to how the CSR performed when she/he answered the call. The evaluator also marked the remaining eight questions as 0 since they are not applicable to the call. The final score of customer service satisfaction is calculated by the sum of the score of all applicable questions divided by the number of applicable questions. In this example, the score is 8, the sum of 3, 4 and 1, divided by 3 applicable questions, which is 2.67.

Six calls are randomly selected for each CSR in each month. The monthly score is the total score of applicable questions of all six monitored calls divided by the total number of applicable questions in all six monitored calls.

Business need satisfaction is evaluated the same way as customer service satisfaction. However, the questions are varying from one product to another. Typical questions of business need satisfaction are "did a CSR provide correct information to customer" or "did a CSR access proper systems or documents". The total number of questions varies too. The minimum number of the questions is eight. The max number is sixteen. The way to score the business needs is exactly the same as customer service satisfaction.

Although the final scores of customer service satisfaction and business need satisfaction are continuous numbers ranging from 1 to 5, in the call centers, which were the source of the data used in the research, these results are converted to monthly evaluations according to the rules in Table 1.

Table 1: Quality Evaluation Converting Scales

Not Met	Score < 2
Met Some	Score >= 2 and Score < 3
Met	Score >= 3 and Score < 4
Exceeded	Score >= 4 and Score < 4.75
Far Exceeded	Score >= 4.75

4.1.2 Time Management

The attributes of time management are *adherence, after call work time, aux and attendance.* The data of time management are collected from phone switches on monthly basis.

Adherence is the percentage of the length of time a CSR is logged into the phone switch to the length of time he/she is supposed to be logged in. *After call work time* is the average number of seconds that a CSR spend on processing works after the calls during a month. *Auxiliary* is the percentage of the length of time a CSR is spending on personal activity to the length of time that a CSR is logged into the phone switch. *Attendance* is a CSR's monthly absence.

Finally in the available data, there is a Boolean attribute representing the fact that the CSR is in a training period. Each record has a time stamp. There is an attribute representing which product a CSR is serving in. In summary there are total ten attributes in the dataset and they are presented in table 2.

Table 2: Dataset Description

Category	Attribute Name	Data Type	Format	Example
	Agent ID	Integer		1, 201, etc
	Date of Data	Date	mm/01/yyyy	09/01/2001
	Training	Boolean	0, 1	0
	Product ID	Integer		226, 3927
Quality	Customer Service	Category	1, 2, 3, 4	3
	Business Needs	Category	1, 2, 3, 4, 5	4
Time management	After Call Work Time	Integer	1, 2, 3,	180
	Adherence	Float	Percentage	96%
	Attendance	Integer	1, 2, 3, ...	2
	Auxiliary	Float	Percentage	4%

4.2 Data Cleaning and Preparation

The values of customer service satisfaction and business needs should all fall between one and five. The data below one represents missing. The data above five were not found in the dataset. After converting to categories, the record values of customer service satisfaction should have only 1, 2, 3, 4 and business need satisfactions should have only 1, 2, 3, 4, 5.

The distribution of customer service satisfaction did not have good distribution. Only six records fell into not met and thirteen fell into far exceeded. The records fell into not met and far exceeded were deleted from the used dataset since the data is too few to give enough training information and nor to enough records for the testing. The majority of the records fell into met class. This met class is separated to two sub-classes at 3.5. The research trained both big met class and sub-class and compared performance of built models.

The value of time management categories should all be equal to or above zero. The values below zero are not valid and have to be deleted. The records that have missing values are also deleted from the final dataset.

After cleaning, total 14671 records are in the final customer service dataset. Total 14690 records are in the final business need dataset.

The distributions of the final dataset for customer service satisfaction are as following:

Table 3: Distribution of Customer Service Satisfaction

Customer Service Satisfaction			
Class 1 – Met Some	>=2 and < 3	1469	10.01%
Class 2 – Met 1	>=3 and < 3.5	5965	40.66%
Class 3 – Met 2	>=3.5 and < 4	5841	39.81%
Class 4 - Exceeded	>=4 and < 4.75	1396	9.52%
	Total	**14671**	100.00%

Table 4: Distribution of Customer Service Satisfaction without Splitting Met Class

Customer Service Satisfaction			
Class 1 – Met Some	>=2 and < 3	1469	10.01%
Class 2 – Met	>=3.5 and < 4	11806	90.47%
Class 3 - Exceeded	>=4 and < 4.75	1396	9.52%
	Total	**14671**	100.00%

The distributions of the final dataset for business need satisfaction are as following:

Table 5: Distribution of Business Need Satisfaction

Business Need Satisfactions			
Class 1 – Not met	>=1 and < 2	63	0.43%
Class 2 – Met Some	>=2 and < 3	3533	24.05%
Class 3 – Met	>=3 and < 4	5974	40.67%
Class 4 – Exceeded	>=4 and < 4.75	3610	24.57%
Class 5 – Far Exceeded	>=4.75 and <=5	1510	10.28%
	Total	**14690**	100.00%

Different products have different expectation to CSRs on after call work, adherence and auxiliary. For example, 150 seconds may be a short after call work time for one product but be too long for another product. So after call work time, adherence and auxiliary are scaled to float numbers from 0 to 1 before the records of different products are put together according to following formula. X represents a value of a cell, min(a, p) represents the minimum value of an attribute of a product, and max(a, p) represents the maximum value of an attribute of a product.

$$(x - min(a, p)) / (max(a, p) - min(a, p))$$

After all the records from different products are put together, other attributes except date are further scaled to float number from 0 to 1.

There are eight input attributes in the final dataset, which are agent ID, date, product ID, training, ACW, aux, adherence and attendance. The output attributes are two, which are customer service satisfaction and business needs. To achieve the best performance, one network is trained for each class. There are four classes for customer service satisfaction and five classes for business needs.

4.3 Methodology

In the research we used three software packages including Trajan, WebStat and SVM-light to mine the performance data. Trajan is a neural network simulation package developed by the Trajan Software Ltd. Trajan supports a large number of neural network architectures. Trajan has powerful variable selection. Its design process is near automatic. Trajan delivers fast training non-linear model [37]. WebStat is a web based data mining and statistics tool. It includes most data mining algorithms, statistics method and neural networks. The research focused our deployment on Multi-Layer Perceptron, Probabilistic Neural Networks, Linear Models, CART, SVM and Decision Tree-ANN Hybrid Models. The research compared the performance of the six algorithms.

All the algorithms use random sampling. Each class is trained by using each algorithm several times. The results from same algorithm are close so we could make the assumption that the results are representative.

Below we will introduce the process of how the models are set up in more details.

4.3.1 Multi-Layer Perceptron

Multi-layer perceptron is the most popular neural network architecture. It consists of at least three layers, an input layer of source neurons, at least one hidden layer of computational neurons, and an output layer of computational neuron [23]. The input layer accepts inputs and redistributes to all the neurons of the middle layer. The neurons in the middle layer detect the features of input patterns and pass the features to the output layer. The output layer uses the features to determine the output patterns.

In out research we use one hidden layer. The default number of neurons in the hidden layer is a hundred thirteen for both business need and customer service satisfaction. After deploying the neuron numbers from fifty to a hundred-twenty, the research found that the results by using different number of neurons are very similar. A hundred thirteen is adopted since the performance seems better than others.

There are eight neurons in the input layer since there are eight input attributes. There is one neuron in each model for one output. The network is illustrated in figure 1.

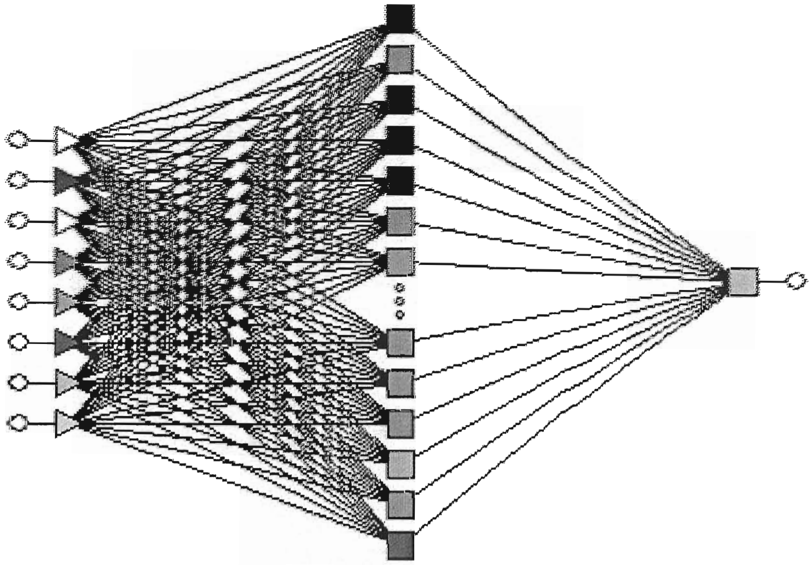

Figure 1: MLP Model

MLP training may be a one or a two-phase training. Back propagation alone is used in the case of one phase training. In the case of two-phase training back propagation is used first and then the conjugate gradient is used. Conjugate gradient serves as a second phase algorithm that it keeps zero second derivative on the assumption that the surface is nice smooth [37].

The research uses random sampling during the first phase training. Trajan assigns 50% to training process, 25% for testing, and the remaining 25% for out-side testing or validation. The second phase uses the same sampling that the first phase used for comparison purpose.

During the experiments, we used a hundred epochs for both back propagation and conjugate gradient since Trajan restored to the best network that is using less than sixty epochs. Sixty is chosen for conjugate gradient since the training error remains the same soon after the second phase is started.

4.3.2 Linear Networks

We used linear neural networks available in Trajan. Linear networks have two layers: an input layer and an output layer. Linear models have good performance on linear problems. However, they cannot solve more complex problems. Linear networks can be trained to serve as a standard comparison for non-linear problem [37].

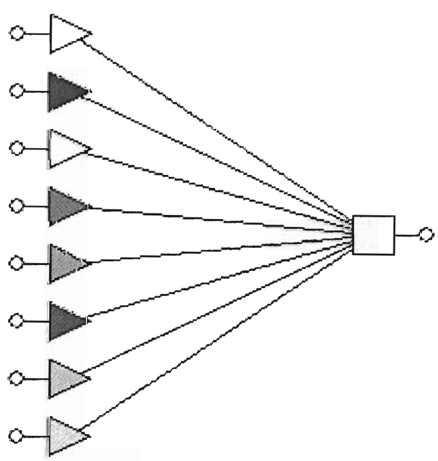

Figure 2: Linear Model

Linear model is relatively simple and no parameters need to be selected by the users. The algorithm available in Trajan is the standard pseudo-inverse (SVD) linear optimization algorithm.

4.3.3 Probabilistic Neural Networks

Probabilistic Neural Networks have been developed only for classification problems. PNN are kernel-based estimation. They usually have three layers: one input layer, one hidden layer and one output layer. The network copies the training cases into the hidden layer. The hidden layer has always the same number of cases. The output layer adds the estimates and produces the output. In this research, there are 7337 neurons for training the customer service attribute and 7346 for business needs attribute in the hidden layer. Prior is the weights setting of "Correct" class and "Wrong" class. The research did not use prior since it yields very low accuracy.

The Probabilistic Neural Networks can be illustrated as figure 3.

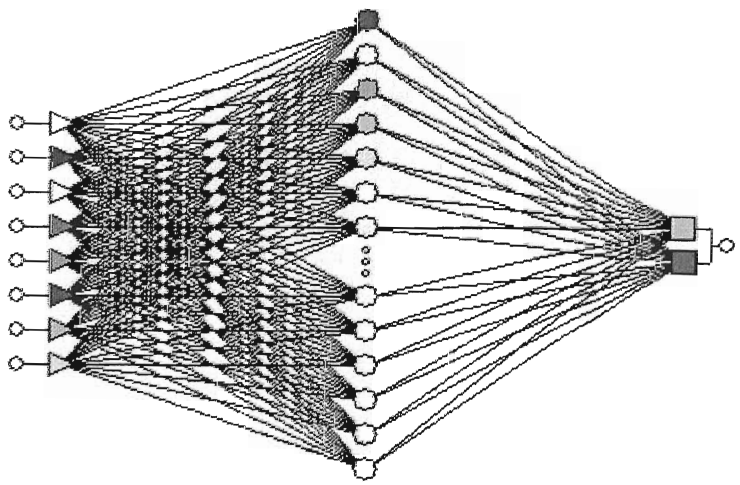

Figure 3: PNN Model

4.3.4 Advanced Classification Trees (CART)

The research uses advanced CART available in WebStat. Decision tree mode is a tree structured binary decision. Each decision tree has internal nodes and leaf nodes. Leaf nodes represent the final decision or prediction. CART labels each leaf node a unique increasing integer number from left to right starting from 1. All the records in the dataset are assigned an integer.

CART creates decision trees to predict categorical dependent by using both categorical and continuous predictors. Advanced CART optimizes the options for large and very large dataset.

Gini was selected for goodness of fit measurement to achieve the best performance. The maximum number of tree level that can be used in WebStat is 32. The research uses 32 as maximum tree height since the tool can restore to the tree structure that has best performance. The training selects 10-fold cross validation sampling to achieve best training and testing.

A glance of a sample constructed tree layout is as figure 4.

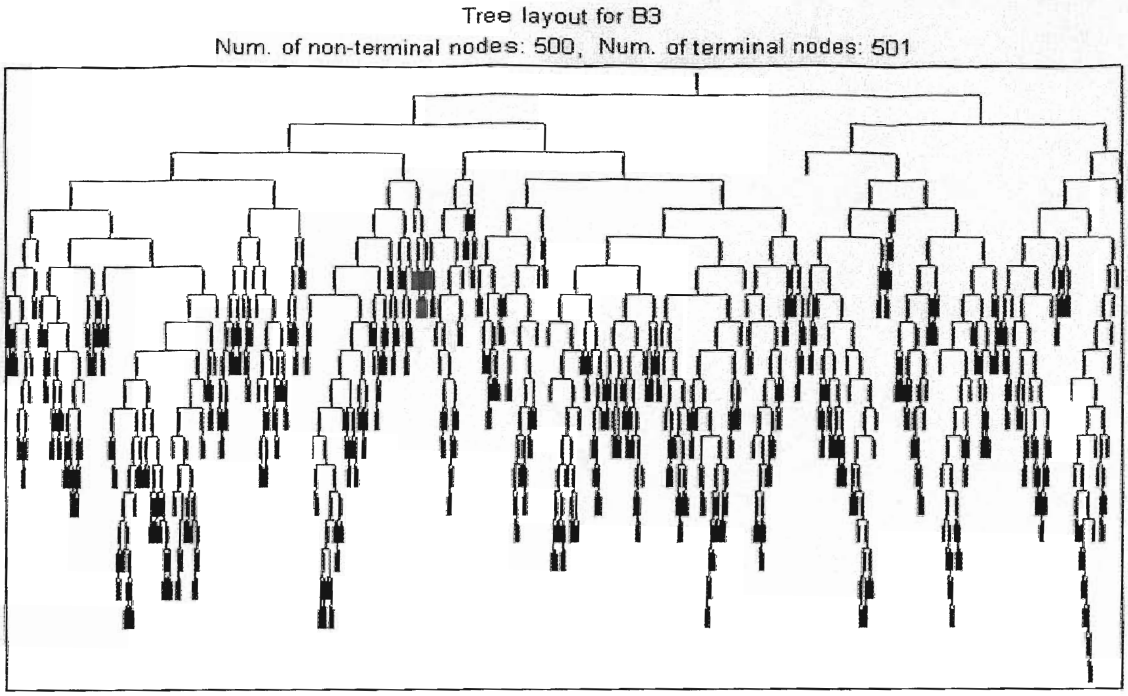

Figure 4: CART Model – Tree Layout
B3: Met class of business

4.3.5 Decision Tree - ANN Hybrid Model

The idea of Decision Tree – ANN Hybrid Model is that the data is fed into decision tree first. Then the leave notes information is obtained and added into the dataset. Artificial Neural Networks (ANN) training and testing are performed on this new dataset

We used CART available to build the decision tree. The leaf node labels, which were unique integers as discussed in session 4.3.4, were obtained after CART built the tree. A new attribute of node label was added into the original dataset.

We used back propagation for the artificial neural network model building. The parameters we used are the same as the ones for CART training alone and BP alone. The final model is as figure 5.

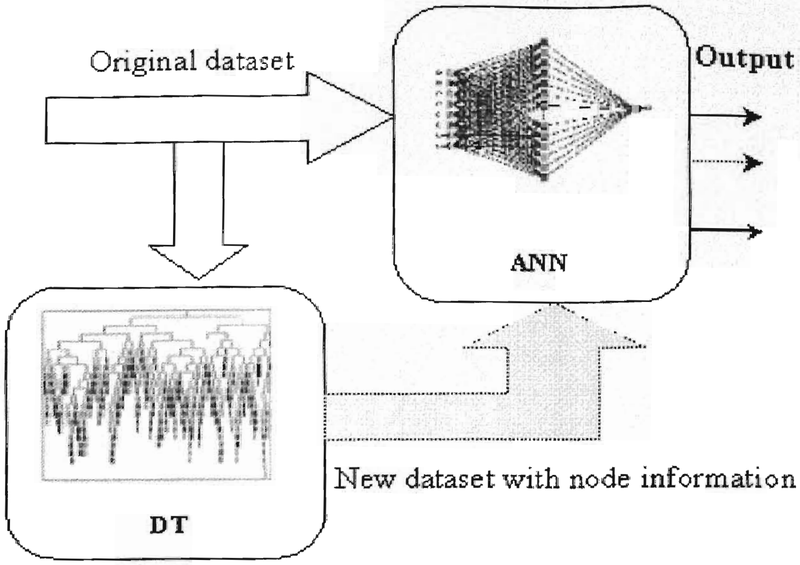

Figure 5: Decision Tree-ANN Hybrid Model

4.3.6 Support Vector Machine

Support Vector Machine (SVM) is a learning method introduced by V. Vapnik et al [40]. SVMs perform binary learning. They only distinguish positive and negative two classes. It conducts computational learning based on *Structural Risk Minimization*. Structural risk minimization finds a hypothesis h for which the lowest true error is guaranteed. The true error of h is the probability that h will make an error on an unseen and randomly selected case. An upper bound of the true error can be used for h. Support vector machine finds the hypothesis h and minimizes the bound of the true error [41].

The research uses SVM_light package developed by Thorsten Joachims. The research experienced one to six degrees by using polynomial function. We found that the model with degree 3 has best performance.

CHAPTER V

RESULTS ANALYSIS

In this chapter, we analyzed he results from two perspectives: the performance of each model and the input importance that each model used.

5.1 Performance Analysis

The performance measurement is calculated from confusion matrix of testing results. Twenty-five percent data were used form cross-validation testing in linear model, MLP, PNN and SVM. The 10-fold cross-validation testing was done for CART. The performance result is the sum of total number correct prediction of "Correct" category and the correct prediction of "Wrong" category divided by the total number of testing case.

The performance of a perfect model is 100% for both "Correct" category and "Wrong" category. The models that have accuracy near to 100% are good. A random classifier should have 50% accuracy.

Table 6 shows the performance of each model for predicting customer service satisfaction. The results of met class are shown in smaller font as a comparison of separated sub-classes. According to the overall results from the confusion matrix, the ranking of the performance of the trained models is CART, PNN, SVM, BP/CG, BP, Hybrid and the Linear Models. There are no apparent difference among BP/CG, BP and Hybrid.

Table 6: Classification Accuracy of Customer Service Prediction

Customer Service Skills – Cross Validation										
Class		Case #	Linear	BP	CG	BP/CG	PNN	CART	Hybrid	SVM
Met Some	Correct	1469	68.77	66.77	60.28	68.88	0.00	90.13	66.96	0.00
	Wrong	13202	66.67	70.71	58.91	70.68	100.00	83.08	70.47	100.00
	Overall		68.56	70.38	59.04	70.52	90.26	91.65	70.33	89.95
Met 1	Correct	5969	58.16	60.29	54.35	60.80	28.78	74.43	62.80	18.44
	Wrong	8702	60.31	61.24	54.77	60.73	86.37	70.37	58.66	90.64
	Overall		59.04	60.87	54.60	60.76	63.13	74.65	60.40	61.28
Met 2	Correct	5841	59.40	59.15	51.25	60.12	34.63	83.79	61.07	22.79
	Wrong	8830	59.93	61.75	52.85	62.88	81.55	63.59	61.95	88.65
	Overall		59.72	60.73	52.22	61.79	64.93	73.85	61.60	62.54
Met(1 and 2)	Correct	11810	55.77	61.29	47.46	60.87	99.79	74.69	61.88	100.00
	Wrong	2861	54.30	62.98	45.44	62.81	0.35	83.94	61.45	0.00
	Overall		55.49	61.62	47.07	61.25	89.57	76.50	61.61	80.30
Exceeded	Correct	1396	65.58	67.25	50.29	68.71	0.00	91.12	65.08	0.00
	Wrong	13275	63.32	68.51	49.14	68.72	100.00	84.12	71.43	100.00
	Overall		65.37	68.39	49.25	68.72	90.97	82.36	70.85	50.35

For example for Met 1 class, there were 5969 records out of 14671 falling into "Correct" category in the dataset and the remaining 8702 records fell into "Wrong" category. CART predicted 4443 out of 5969 correctly, which was 74.43% shown as correct prediction of "Correct" class. CART predicted 6124 out of 8702 correctly, which was 70.37% shown in Table 6.

Since 25% of the records in the dataset were used for cross validation for linear, MLP, PNN, and SVM, which is different from CART, the base to calculate the accuracy was different from CART, which was 3668. For example for met 1 class again, 1448 records out of 3668 fell into "Correct" category and the remaining 2220 records fell into "Wrong" category. 873 records out of 1448 were predicted as "Correct" correctly which is 60.29%. 1359 out of 2220 were predicted as "Wrong" correctly, which is 61.21% shown in Table 6.

Table 6 also shows the accuracy details for customer service satisfaction. The research predicted met class and also predicted each met sub-class by splitting the met class into two. Usually the prediction of one large class has higher accuracy. However, it is not true for met class of customer service satisfaction. The big class has more noise. The performance of predicting one large class is very close to the performance of predicting sub-classes. According to this fact, the research reveals that the scale used for customer service evaluation is not proper and caused data without good differentiation. The CSRs in sub-class 1 are more likely met-some performers. The CSRs in sub-class two are more likely exceeded performers.

Table 7 shows the performance of each model for predicting business need satisfaction. The way to calculate the performance of business need prediction is exactly the same as the way for customer service. The ranking of the performance is the same as the models for customer service.

After looking into the detail accuracy of each correct/wrong class, the research found that PNN models are not valid for the dataset used. The research trained several PNN by using different smooth factors, with or without setting priors and also trained with different setting of prior. Priors are the weights of "Correct" category and "Wrong" category. The accuracy is either too low for "Correct" or too low for "Wrong". SVM has the same problem as PNN and the model built by using SVM is not valid either.

Table 7: Classification Accuracy of Business Need Prediction

Business need Satisfaction - Cross Validation										
Class		Case #	Linear %	BP %	CG %	BP/CG %	PNN %	CART %	Hybrid %	SVM %
Not met	Correct	63	50.00	53.85	53.85	53.85	0.00	100.00	65.00	0.00
	Wrong	14608	74.80	80.24	65.70	81.91	99.97	96.45	87.92	100.00
	Overall		74.73	80.15	65.66	81.81	99.46	99.62	87.80	99.73
Met Some	Correct	3533	76.63	80.29	43.24	79.05	52.77	93.43	91.32	57.96
	Wrong	11138	75.33	81.14	40.66	81.90	91.73	83.38	82.63	90.59
	Overall		76.33	80.94	41.29	81.21	82.52	89.14	82.33	82.98
Met	Correct	5974	66.14	70.36	62.35	70.23	52.20	82.64	71.02	50.79
	Wrong	8697	60.30	68.03	59.38	67.94	81.40	75.03	69.07	90.59
	Overall		62.67	68.98	60.59	68.87	69.53	79.82	69.88	69.84
Exceeded	Correct	3610	68.31	73.77	55.77	74.22	23.10	93.82	76.52	24.57
	Wrong	11061	72.46	74.78	50.09	75.77	94.23	79.71	73.93	94.74
	Overall		71.46	74.54	51.53	75.38	77.12	86.51	74.59	76.75
Far Exceeded	Correct	1510	71.03	74.92	59.22	75.83	2.12	96.82	78.00	0.00
	Wrong	13161	75.68	78.78	58.70	79.32	99.46	85.81	82.73	100.00
	Overall		75.17	78.43	58.74	79.00	90.69	92.33	80.84	96.12

BP/CG models are supposed to have better performance than BP alone according to purpose of CG [37]. The typical training chart is showing as figure 6. According to the figures that showing in above tables, the performance of conjugate gradient alone has poor performance. The performance of BP/CG is a bit better than BP. However the results are very close. It is not proper to make the conclusion that the models trained by BP/CG have better performance than the ones from BP alone based this research.

The performance of hybrid model was expected at least the same as CART. However, the final accuracy from this research is a bit better than BP and BP/CG model.

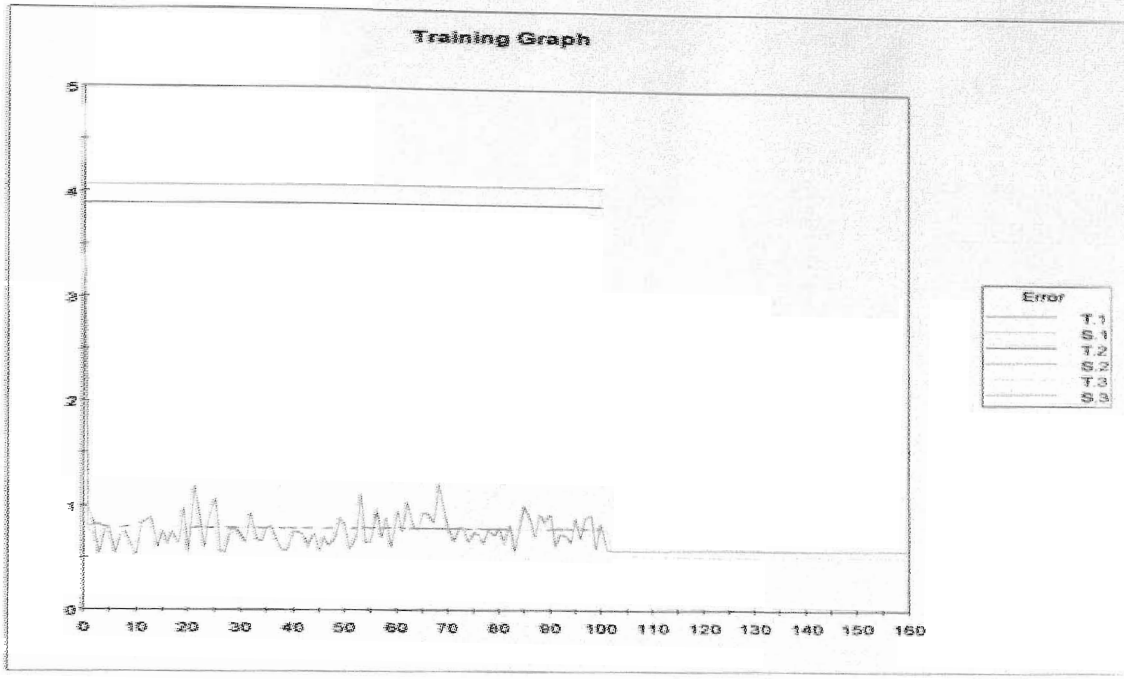

Figure 6: Training Chart of Exceeded Class of Business Needs
T1: Training errors of back propagation
S1: Cross Validation errors of back propagation
T2: Training errors of conjugate gradient
S2: Cross Validation errors of back propagation/conjugate gradient
T3: Training errors of back propagation/conjugate gradient
S3: Cross Validation errors of back propagation/conjugate gradient

The Linear model serves as a comparison for other models. The models trained by other algorithms are supposed to have at least the performance that linear models can get. The performance of linear models is better than the performance of CG only after getting rid of PNN models.

CART models have the best performance in the research. They not only have the best overall performance, but also they have highest accuracy to predict "Correct" (C1) and "Wrong" (C0) for all each class. Figure 7 shows the confusion matrix chart of met class from business needs attributes.

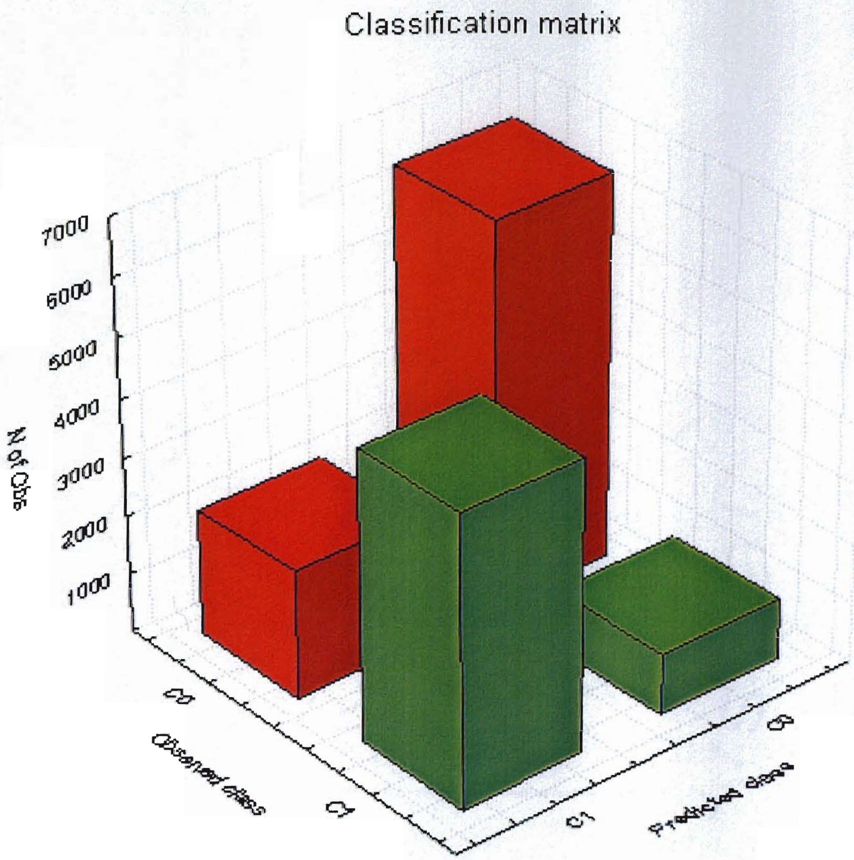

Figure 7: Confusion Matrix of Met Class of Business Needs.
Plot C1-C1 (high green bar): correct prediction of "Correct" category
Plot C0-C0 (high red bar): correct prediction of "Wrong" category
Plot C1-C0 (low green bar): incorrect prediction of "Correct" category
Plot C0-C1 (low red bar): incorrect prediction of "Wrong" category

5.2 Inputs Sensitivity Analysis

The sensitivity analysis is available in Trajan but not available in WebStat. Trajan allows the prediction conducted when there is one or more attributes in absent. The sensitivity is calculated by the accumulated errors when a particular attribute is removed from the training. When an attribute is removed from the training model, the higher the error is, the more important the attribute is. The importance of the inputs is ranked by the accumulated error.

PNN analysis and SVM analysis are not included since the models are not valid. CG is not included since the performance is not good and the accuracy is near random accuracy, which is 50%.

Table 8: Inputs Importance Ranking of Customer Service

Customer Service Satisfaction - Sensitivity Analysis										
Class	**Algorithms**	**Agent**	**Date**	**Training**	**Product**	**ACW**	**Adherence**	**Aux**	**Attendance**	**Note**
Met Some	Linear	7	1	5	3	4	2	8	6	
	BP	3	1	2	4	6	7	8	5	
	BP/CG	8	1	3	2	7	5	6	4	
	Hybrid	4	1	8	3	9	6	5	7	2
Met 1	Linear	3	1	5	4	6	8	2	7	
	BP	2	7	6	1	8	4	3	5	
	BP/CG	2	8	6	1	5	3	7	4	
	Hybrid	2	3	5	1	8	4	6	7	9
Met 2	Linear	2	1	5	7	8	3	4	6	
	BP	8	6	7	1	3	2	5	4	
	BP/CG	4	2	5	1	7	3	8	6	
	Hybrid	8	5	9	6	3	2	7	4	1
Met (1 & 2)	Linear	4	1	8	5	2	3	6	7	
	BP	8	1	3	7	6	5	4	2	
	BP/CG	8	1	3	7	6	5	4	2	
	Hybrid	7	1	4	6	3	5	8	9	2
Exceeded	Linear	2	1	7	3	5	6	8	4	
	BP	7	1	5	2	4	3	6	8	
	BP/CG	8	1	2	4	5	3	6	7	
	Hybrid	4	3	6	1	9	8	2	7	5

Table 8 is the sensitivity ranking of customer service prediction. Product is very sensitive on predicting customer service satisfaction, which indicates that CSRs in some products have more opportunity to far exceed than the CSRs in some other product. Adherence is important too. Adherence is how much a CSR want to adhere to his/her job.

Adherence reveals the attitude toward work. A good attitude may lead to good customer service performance.

Another interesting characteristic is that the date is important when predicting customer service satisfaction. The reason why date is important maybe that the dates are affected by call types. One type of calls may be dominant of all types of calls during a certain period. After that period, calls of another type become the majority in the call volume in next period. Call types are not tracked in this research. The affect of call types may appear as date. Another fact can be training or coaching delivery date. The customer service satisfaction may be improved right after the coaching or training delivery and may be dropped after a certain period of the delivery.

The ranking analysis from Linear Models, BP models, BP/CG models and Hybrid models are pretty consisting on predicting business need satisfaction. The input sensitivity ranking of business need satisfaction is shown in Table 9.

The product become more important on predicting business needs from not met class to far-exceeded class. This can be interpreted that a CSR has more opportunity to be far exceeded if a CSR is in some product. A CSR has less opportunity if he/she is in some other product.

Agent is more important when predicting exceeded and far-exceeded classes. It means that the top performers are likely staying on the top most time. The performance of the CSRs whose performance fall into met or below met is not stable. However, they are more likely staying in met class or below.

Table 9: Inputs Importance Ranking of Business Needs

Business Need Requirements - Sensitivity Analysis										
Class	**Algorithms**	**Agent**	**Date**	**Training**	**Product**	**ACW**	**Adherence**	**Aux**	**Attendance**	**Note**
Not Met	Linear	2	7	1	6	4	8	5	3	
	BP	7	1	2	6	4	8	3	5	
	BP/CG	7	1	6	8	4	5	2	3	
	Hybrid	4	5	9	2	8	7	3	6	1
Met Some	Linear	6	2	5	3	4	1	8	7	
	BP	4	3	7	1	8	2	5	6	
	BP/CG	4	3	5	1	8	2	6	7	
	Hybrid	4	2	3	1	8	9	5	6	7
Met	Linear	2	1	6	4	3	5	7	8	
	BP	7	1	4	2	8	3	5	6	
	BP/CG	8	1	6	2	7	3	5	4	
	Hybrid	7	3	6	2	8	4	5	9	1
Exceeded	Linear	6	8	4	3	2	1	5	7	
	BP	3	6	8	2	4	1	5	7	
	BP/CG	5	3	7	2	4	1	6	8	
	Hybrid	5	2	8	9	4	1	3	7	6
Far Exceeded	Linear	2	3	7	1	6	5	4	8	
	BP	2	4	8	1	7	3	5	6	
	BP/CG	2	4	5	1	8	3	6	7	
	Hybrid	3	4	7	1	8	6	5	9	2

CHAPTER VI

CONCLUSION

The research built six types of models to predict quality score of customer service satisfaction and business need satisfaction by using Linear models, Multi-Layer Perceptrons, Probabilistic Neural Networks, CART, Decision tree-ANN Hybrid model and SVM. The research compared the performance of the six types of models based on the confusion matrix results of cross validation. The performance is also analyzed by using the accuracy of "Correct" category prediction and the accuracy of "Wrong" category prediction. The winning models are the ones that were trained by using CART in WebStat. The overall accuracy from CART is 80.63% on predicting customer service satisfaction and 89.48% on predicting business need satisfaction. The accuracy of "Correct" category and the accuracy of "Wrong" category are very close. The trained models based on CART can be used for future prediction. Back propagation with conjugate gradient did not have significant better performance than back propagation only.

The research also analyzed the sensitivity of inputs. The research found that products, agents and dates could affect the quality of performance than time management. The CSRs serving in some products have more opportunity to exceed the expectation than the ones in some other products. The top performers constantly exceed or far-exceed the expectation. The performance of CSRs whose evaluation results falling into met or below is not so stable as top performers. The research suggest that call center management team should focus training and coaching the individuals and products that constantly have low quality instead of emphasizing balancing the length of times spent on calls.

REFERENCE

1. Altitude Software, Improving Agent Performance While Maintaining High Level of Motivation, http://www.realmarket.com/required/altitude1.pdf, September 2002, access date: April 2003

2. Eric Bloedorn, Alan D. Christiansen, William Hill, Clement Skorupka, Lisa M. Talbot, Jonathan Tivel, Data Mining for Network Intrusion Detection: How to Get Started, http://www.mitre.org/support/papers/tech_papers_01/bloedorn_datamining/bloedorn_datamining.pdf, access date: September 2002

3. James Brewton, Customer Value-Driven Competency Models: Powerful Tools for Maximizing Call Center Performance and Customer Loyalty, http://www.crm2day.com, 2002, access date: April 2003

4. Susan Bridges and RayFord Vaughn, Intrusion Detection Via Fuzzy Data Mining, Proceedings of the 12th Annual Canadian Information Technology Security Symposium, Ottawa, Canada, June 19-23, 2000

5. A.G. Buchner, M. Baumgarten, S.S. Anand, M.D. Mulvenna, and J. G. Hughes, Navigation Pattern Discovery from Internet Data, http://www.acm.org/sigkdd/proceedings/webkdd99/papers/paper13-buchner.pdf, 2000, access date: September 2002

6. Alex Buchner, Maurice Mulvenna, Discovering Internet Marketing Intelligence through Online Analytical Web Usage Mining, http://www.acm.org/sigs/sigmod/disc/disc99/disc/record/issues/9812/SPECIAL/buchner.pdf, 1998, access date: September 2002

7. Stephan Busemann, Sven Schmeier, Roman G.Arens, Message Classification in the Call Center, *Proceedings of the 6th Conference on Applied Natural Language Processing,* Seattle, WA, 2000

8. Ming-Syan Chen, Jong Soo park, and Philip Yu, Data Mining for Path Traversal Patterns in a Web Environment, http://www.datamining.org.tw/doc/similar/Data%20mining%20for%20path.pdf, 1996, access date: September 2002

9. Srivastava, J., Cooley R., Deshpande M. and Tan P.N., Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data, SIGKDD Explorations.

10. Web mining: Information and Pattern Discovery on the World Wide Web, 9th International Conference on Tools with Artificial Intelligence (ICTAI '97), November 03 - 08, 1997, Newport Beach, CA

11. Robert Cooley, Bamshad Mobasher, and Jaideep Srivastava, Data Preparation for Mining World Wide Web, www.cs.umn.edu/research/websift/papers/kais99.ps, 1999, access date, October 2002

12. Russell Deathridge, Performance Related Pay in the Contact Center, http://www.crm2day.com, 2002, access date: April 2003

13. Linda Dilauro, What's Next in Monitoring Technology? Data Mining Finds a Calling in Call Centers, Dictaphone Corporation

14. Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth. From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, Fall1996, pp. 37-54.

15. A. Gilmore, Call Center Management: Is Service Quality a Priority, Managing Service Quality, vol. 11, no. 3 pp. 153-159

16. HigherGroud, Discovering the Business Intelligence Hidden in Your Call Center, HigherGroud White Paper Document, June 2002, access date: April 2003,

17. Anupam Joshi, Karuna Joshi, On Mining Web Access Logs, http://www.enterpriseware.net/EWRoot/Files/dmkd00.pdf, 1999, access date: October, 2002

18. Karuna Pande Joshi, Analysis of Data Mining Algorithms, http://userpages.umbc.edu/~kjoshi1/data-mine/proj_rpt.htm#Introduction, 1997, access date: September 2002

19. Ro King, Data Mining and CRM, http://www.crm2day.com, 2002, access date: April 2002

20. Raymond Kosala, Hendrik Blockeel, Web Mining Research: A Survey, SIGKDD Explorations: Nesletter of ACM SIG on Knowledge Discovery & Data Mining, June 2000, Volume2, Issue 1

21. Pawan Lingras, Rough Set Clustering for Web Mining, http://citeseer.nj.nec.com/cache/papers/cs/26234/, 2002, access date: September 2002

22. Lizhen Liu, JunJie Chen, Hantao Song, The Research of Web Mining

23. Michael Nigneritsky, Artificial Intelligence: A guide to Intelligent Systems, Pearson Education Limited, Addison-Wesly, 2002

24. Olfa Nasraoui, Hichem Frigui, Anupam Joshi, Mining Web Access Logs Using Relational Competitive Fuzzy Clustering, *In Proceedings of the Eight International Fuzzy Systems Association World Congress*,

http://homepages.wmich.edu/~w0low/mining-web-access-logs.pdf August 1999, access date: September 2002

25. Amy Neustein, Building Natural Language Intelligence into Voice-Based Applications, http://www.speechtechmag.com/issues/7_4/cover/915-1.html, July 2002, access date: April 2003

26. Pal S.K., Talwar V. and Mitra P., Web Mining in Soft Computing Framework: Relevance, State of the Art and Future Directions, IEEE Transactions on Neural Networks, 2000

27. Jia-Yu Pan, FastCARS: Fast, Correlation-Aware Sampling for Network Data Mining, In *Proceedings of IEEE GlobeCOM 2002 - Global Internet Symposium*, 2002. Taipei, Taiwan, November 17-21, 2002 http://www.db.cs.cmu.edu/Pubs/Lib/globecom02fastcars/globecom02fastcars_pdf, access date: October 2002

28. A. Parasuraman, Service Quality and Productivity: a Synergistic Perspective, Managing Service Quality, vol. 12, no. 1, pp. 6-9

29. Jose Pena and Ernestina Menasalvas, Towards Flexibility in a Distributed Data Mining Framework, http://www.cs.cornell.edu/johannes/dmkd2001-papers/p11_pena.pdf

30. CyrusShahabi, Armir Zarkesh, Jafar Adibi, and Vishal Shah, Knowledge Discovery from Users Web-Page Navigation, 7th International Workshop on Research Issues in Data Engineering (RIDE '97) High Performance Database Management for Large-Scale Applications, April 07 - 08, 1997

31. Colin Shearer, The CRISP-DM Model: The New Blueprint for Data Mining, Journal of Data Warehousing, Vol.5, No. 4, Fall 2000, pp.13-22

32. David Sims, What Is CRM, http://www.crmguru.com, 2000, access date: April 2003

33. Warren Staples, John Dalrymple, Rhonda Bryar, Accessing Call Center Quality using the SERVQUAL Model, http://www.cmqr.rmit.edu.au/publications/wsjdrbicit02.pdf, 2002, access date: April 2003

34. Sun Microsystems, Inc, Mining the Value in CRM Data: A CFO Perspective, http://www.crm2day.com/data_mining, 2002, access date: April 2003

35. Pan-Ning Tan, Vipin Kumar, Mining Indirect Associations In Web Data, WebKDD 2001: Mining Log Data Across All Customer Touch Points

36. Pan-Ning Tan, Hannah Blau, Steve Harp, Robert Goldman, Textual Data Mining of Service Center Call Records, KDD 2000

37. Trajan Neural Networks, Trajan Software Ltd. 2001,

38. Graham J. Williams and Zhexue Huang, A Case Study in Knowledge Acquisition for Insurance Risk Assessment using a KDD Methodology, *Pacific Rim Knowledge Acquisition Workshop (PKAW96), October 1996, Sydney, Australia*

39. Osmar Zaiane, Man Xin, Jiawei Han, Discovering Web Access Patterns and Trends by Applying OLAP and Data Mining Technology on Web Logs, IEEE Advances in Digital Liraries Conference, April 22 - 24, 1998

40. Vladimir Vapnik, The Nature of Statistical Learning Theory. Springer, New York, 1995

41. Thorsten Joachims, Text Categorization with Support Vector Machines: Learning with Many Relevant Features. Proceedings of the European Conference on Machine Learning, Springer, 1998.

APPENDIX A: A Sample of Original Data

Agent	Date	Training	Product	Attendance	ACW	Aux	Adherence	Service	Business
715	01-Dec-01	0	17	0	10	0.1103	0.9709	3.21	3
1870	01-Dec-01	0	17	0	8	0.0299	0.9834	3.87	3
1867	01-Dec-01	0	17	0	11	0.1261	0.9723	3.43	3.12
1232	01-Jan-01	0	20	0	270	0.2355	0.92	3	3
5614	01-Jan-01	0	20	0	340	0.1955	0.92	3	3
5642	01-Jan-01	0	20	0	147	0.3761	0.92	3	3
295	01-Jan-01	0	20	0	415	0.316	0.92	2.79	2.64
303	01-Jan-01	0	20	0	257	0.3037	0.92	3.09	3
508	01-Jan-01	0	20	0	293	0.2417	0.92	2.13	3
494	01-Jan-01	0	20	0	83	0.3754	0.92	3.4	3
1232	01-Feb-01	0	20	0	199	0.2332	0.9267	3	3.06
5614	01-Feb-01	0	20	0	634	0.1264	0.8855	3.05	3

VITA ②

Ruiyuan Guo

Candidate for the Degree of
Master of Science

Thesis: Mining Call Center Performance Data

Major Field: Computer Science

Biographical:

Personal Data: Born in Baotou City, P. R. China, on April 21, 1967, the daughter of Peicheng Guo and Shuxian He.

Education: Graduated from Baotou First High School in Baotou, July 1984. Received Bachelor of Education from Beijing Normal University, Beijing, China in July 1988. Received Master of Education from Beijing Normal University, Beijing, China in July 1991. Completed requirements for the Master of Science degree with a major in Computer Science at Oklahoma State University in December 1999.

Experience: Teacher, Baotou University, Baotou, China from August 1991 to January 1993. Graduate teacher assistant, Computer Science Department, Oklahoma State University from August 1999 to December 1999.