Churn Prediction using Complaints Data

John Hadden, Ashutosh Tiwari, Rajkumar Roy, and Dymtr Ruta

***Abstract*—**The aim of this paper is to identify the most suitable model for churn prediction based on three different techniques. The paper identifies the variables that affect churn in reverence of customer complaints data and provides a comparative analysis of neural networks, regression trees and regression in their capabilities of predicting customer churn.

***Keywords*—**Churn, Neural Networks, Regression, Decision Trees.

I. Introduction

IT has become common knowledge within companies, that their most valuable assets are their existing customers. An environment has emerged where customers are offered a huge choice of service providers to choose from, making winning new customers difficult and costly (Van Den Poel and Lariviere, 2003).

Past research has focused on using demographical data for the purpose of churn prediction, however Wei and Chiu, (2002), have identified several reasons to why this type of data is unsuitable. They point out that using demographical data for the basis of churn prediction creates a churn analysis that is dependent on the customer rather than the contract. It is also suggested that demographic data held by some companies is very limited, restricting the suitability of many existing churn-prediction systems.

In response to these problems Wei and Chiu, (2002) based their churn prediction model on call pattern changes and contractual information. As an alternative to this approach the authors have investigated an approach of the suitability of customer complaints and repairs data for churn prediction. The best variables are identified and neural networks, classification trees and regression are compared for their suitability for churn prediction using this type of data.

II. The Data and Variables

The data used for model development includes information about customer complaints and repair interactions with the company. It was brought to the attention of the author that with a big service provider and infrastructure owner it is not possible to make use of demographic and usage data due to monopoly regulations.

Due to data protection and the sensitivity of the data the variables will be loosely described, however a full discussion on the results obtained from each of the technologies will be given. Three groups of variables have been combined to create the dataset:

John Hadden, Ashutosh Tiwari and Rajkumar Roy are with Enterprise Integration, Cranfield University, Wharley End, Bedfordshire, MK43 0TE.

Dymtr Ruta is with British Telecom, Adastal Park, Ipswich, Suffolk, IP5 3RE.

1. Provisions Data – This data represents estimations that are made by the company regarding the resolution of a complaint or repair.
2. Complaints Data – This data represents the information regarding complaints.
3. Repairs Data – This data represents all information about a fault and repair.

Some of the provisions variables include a count of the total number of appointments that were broken. If the count is greater than one then a promise had previously been broken. The provisions data also includes the number of days by which a promise had been delivered late.

The customer complaints data includes variables related to the type of complaint, duration of the complaint from the initial customer contact to resolution, the number of days the complaint ran over the resolution date, the number of complaints made within a certain time period and if the company refunded any money to the customer as a result of the complaint. Repairs data include variables such as the duration of a repair, how many appointments had been made, how many engineers had visited the site and the type of fault.

The company that provided the data is one of the largest in the world in its domain. It contains huge data warehouses holding information on a wide variety of products and services and millions of customers. The information in this paper does not reflect the actual churn rates of the sponsoring company. It was decided to use a small sample of data for training and validation of the models. The training set was scaled down to 202 customers, approximately 50% of the customers were churners and 50% of the customers were non-churners. The validation set contains a total of 700 customers and has been comprised of 30% churners and 70% non churners. Experiments performed on the technologies also involve training and testing on the same data. It was the understanding of the authors that testing the models with the same data used to train them should result in a very high accuracy and provide a measure of how each technology performs.

Wei and Chiu, (2002) randomly selected variables from their case study data warehouse resulting in a dataset of 1.5% - 2% churners and 98% - 98.5% non-churners. Recognising that this ratio would potentially jeopardise the learning effectiveness of their model and result predictions that favour only the majority decision class (non-churners), resulting in a ‘null’ prediction system, they adopted a multi-classifier class combiner approach as proposed by Chan *et al.*, (1999). Given a set of training instances, this method creates *S* multiple training subsets with a desired class ratio, where the instances having the majority decision class are randomly and evenly partitioned into training subsets. The instances belonging to the minority decision class are replicated across the training subsets. So basically the majority set, in this case being the non-churners, would be sub-divided into maybe 10 subsets then the minority set (the churners) will be included in every subset, resulting in ten datasets with a much higher churn to non-churn ratio. The dataset created by the authors are similar in principle to those created by Wei and Chiu (2002) and Chan *et al.,* (1999), and are sufficient to overcome the skewed distribution problem.

III. The Churn Prediction Models

Many techniques have emerged for the purpose of predicting a required outcome. The methods that are most frequently used in research, therefore the most popular, have been recognised as neural networks, classification trees and regression (Au *et al.*, 2003, Hsieh, 2004, Hwang *et al.*, 2004, Datta *et al.*, 2001, Rosset *et al.*, 2002, Rygielski *et al.*, 2002, Boone and Roehm, 2002, Vellido *et al.*, 1999, Kavzoglu and Mather, 2001, Meyer-Base and Watzel, 1998, Dudoit and Van Der Laan, 2003, Baesens *et al.*, 2004, Bloemer *et al.*, 2002, Lee, 1999, Fong and Cheung, 2004, Ho Ha *et al.*, 2002, Van Den Poel and Lariviere, 2003, Shin and Sohn, 2004). The authors have investigated the suitability of these technologies for predicting customer churn using the complaints data previously mentioned.

A model was created for each of the chosen technologies (Linear Regression, Regression Tree and Neural Network). Each of the models was constructed using the same training dataset. The training dataset comprised of 202 customers with a 50:50 ratio of churners and non-churners.

The regression tree and neural network models were created using Matlab. Matlab is a high-level language and interactive environment that enables its users to perform rigorous tasks much faster than they could be done using conventional programming languages. Matlab has many pre-written toolboxes that can be used with minimum training to perform experiments using many advanced technologies. Some examples of toolboxes available include neural networks, statistics, fuzzy logic and signal processing. The author has used Matlab with the neural network toolbox and the statistical toolbox.

The linear regression model was developed using SPSS. SPSS was used to identify the variables used in the regression model and the regression equation was constructed from these.

A neural network is an analogous data processing structure that possesses the ability to learn. The concept is loosely based on a biological brain and has successfully been applied to many types of problems, such as classification, control, and prediction (Behara *et al*., 2002).

Classification and regression trees (CART) are constructed by recursively splitting the instance space into smaller subgroups, although only until a specified criterion has been met. The decrease in impurity of the parent node against the child nodes defines the goodness of the split. The tree is only allowed to grow until the decrease in impurity falls below a user defined threshold. At this time the node becomes a terminal, or leaf node.

Regression analysis is a popular technique used by the researchers dealing with predicting customer satisfaction. Even those who have aimed to develop more complex models have initially begun their research with basic regression models (Kim and Yoon, 2004).

IV. Neural Networks

Several experiments where performed using feed-forward, back propagation neural networks. The architecture remained the same however different activation functions where applied to assess which activation function performed the best for predicting customer churn based on the authors data. All architectures consisted of twenty four inputs, two layers and the output layer. Various activation functions where also investigated. An illustration of the architecture can be viewed in Fig. 1.

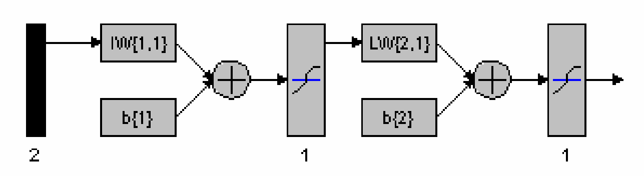

Fig. 1 Neural Network Architecture

First the neural network was constructed using the logsig activation function and trained using the training dataset containing a 50/50 ration of 202 customers. The performance of the neural network was then validated by instructing it to predict the churners from the same training set. Churn is indicated in our dataset as a 0 for non-churn or a 1 for churn, however due to the nature of the neural network a decimal value between 0 and 1 is predicted as for each customer (the churn risk). The greater the churn risk, the more likely that customer is of churning. This allows the authors to manually define the churn threshold and could be viewed as 'fine tuning' the results. The varied churn risks obtained from the neural network is illustrated in Fig. 2.

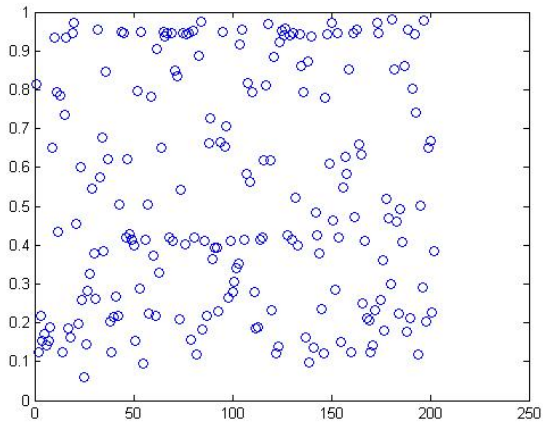

Fig. 2 Logsig Churn risk Predictions

It can be observed from Fig. 2 that the churn risks for each of the customers are extremely varied. Several experiments where performed to establish the most accurate churn threshold. It was established that a threshold of 0.7 provided the most accurate results for predicting customer churn. The confusion matrix in Fig. 3 illustrates the results obtained from basing the neural network on the logsig function and setting the churn threshold to 0.7.

True Labels	Estimated Labels 0	1	Totals
0	54	40	94
1	10	98	108
Totals	64	138	202

Fig. 3 Logsig Function Churn Results

It can be viewed from the confusion matrix in Fig. 3 that the results where reasonably accurate. We will measure the accuracy using a simple formula by dividing the predicted churners by the actual churners. This shows that the feed-forward, back-propagation neural network using the logsig activation function has provided 90% accuracy.

Neural networks where also created with the purelin activation function and Satlin activation function. The purelin activation function also provided 90% accuracy but the Satlin activation function gave 41% accuracy. Based on these results, the authors decided to construct their neural network using the logsig activation function.

With the best activation function identified the authors constructed another neural network using the Bayesian architecture. Both the Bayesian and standard feed-forward back propagation neural networks where trained using the training dataset and validated using the validation dataset. The results are displayed in Fig. 4.

Bayesian Architecture

True Labels	Estimated Labels 0	1	Totals
0	371	63	434
1	119	147	266
Totals	490	210	700

Standard Architecture

True Labels	Estimated Labels 0	1	Totals
0	386	93	479
1	104	117	221
Totals	490	210	700

Fig. 4 Neural Network Validation

As can be viewed in Fig. 4, the Bayesian architecture correctly identified 70% of the churners while the standard feed-forward architecture only identified 55% of the churners. However the standard architecture correctly identified 79% of the non-churners while the Bayesian architecture identified 75% of the non-churners correctly. The real difference between the two architectures can be seen when the actual prediction accuracy is calculated. The prediction accuracy can be calculated by dividing the total number of correct predictions by the size of the dataset. This shows that the Bayesian framework has a prediction accuracy of 74% and the standard framework has a prediction accuracy of 72%.

An analysis of the weights established for the twenty four variables that where presented to the neural network suggests that seven variables held more significance for predicting customer churn than the others. These variables are as follows:

1. How many engineers arrived on site
2. How long the customer had been with the company
3. How long the repair took
4. No. of appointments made for repair
5. The resolution time
6. If an order has been placed
7. How many times the repair has been transferred

V. Decision Trees

Classification trees where used to establish if they would offer a better method of predicting customer churn using complaints and repairs data. The authors chose to use regression trees rather than classification trees because one of the aims of the research is to identify the customer churn risk. The target data is either represented by a 0 for non-churn or a 1 for churn so the classification tree will return a 0 or 1 when making predictions. The authors require the model to return a decimal value.

Twenty four variables where used to train the train the regression tree however the resulting rules where based on only seven of these variables. These variables are as follows:

1. Cash Conceded
2. No. of Complaints
3. No. of appointments made for repair
4. Type of customer
5. No. of missed appointments
6. No. of appointments made to a customer
7. If an order has been placed

The variables selected for the rules are not very surprising and create a fairly accurate rule base for predicting customer churn. When the tree was used to predict churn on the validation dataset the following results where obtained:

True Labels	Estimated Labels 0	1	Totals
0	435	71	506
1	55	139	194
Totals	490	210	700

Fig. 5 Regression Tree Churn Prediction

It can be observed from the confusion matrix in Fig. 6 that the regression tree provided good results for churn prediction correctly predicting 66% of the churners and 88% 0f the non-churners, with an overall accuracy of 82%. Another advantage of using a classification tree is the extracted rules could be used for creating a fuzzy clustering model however that is beyond the scope of the current research but maybe investigated at a later stage.

VI. Regression

The final method that was used for analysis was regression. SPSS was used to calculate the standard error rate for each of the variables used in the regression model and out of the total of twenty four variables only fourteen where used in the model. With the standard error rates calculated we can obtain the seven variables that hold the most significance for predicting customer churn in respect to linear regression. These variables are as follows:

1. No. of appointments made to a customer
2. How many times the repair has been transferred
3. How long the repair took
4. Type of complaint
5. No. of Complaints
6. No. of repairs
7. How long a promise was delivered late

With the variables selected and the standard error for each of these variables determined, the regression model was constructed using the following formula:

$$
\begin{aligned}
Churn = {} & 0.480 + (-0.074)*(Lifetime) + (+6.311)*(P\_Duration) + (-8.715)*(P\_Overdue) \\
& + (-2.314)*(R\_Duration) + (+1.699)*(R\_Promise) + (-0.002)*(R\_MU\_Count) \\
& + (+2.913)*(C\_Duration) + (+1.225)*(C\_Overdue) + (-1.449)*(C\_Issue) \\
& + (+0.098)*(C\_Cash\_conceded) + (+2.835)*(P\_Count) + (+0.617)*(R\_Count) \\
& + (+0.016)*(R\_Overdue) + (-2.589)*(C\_Count)
\end{aligned}
$$

With the regression model created predictions where made on the validation dataset. The results can be viewed in Fig. 6.

```
       | Estimated Labels
 True  |
Labels |    0      1   | Totals
-------|---------------|-------
0      |  465    103   |  568
1      |   25    107   |  132
-------|---------------|-------
Totals |  490    210   |  700
```

Fig. 6 Results for Regression

It can be observed from the confusion matrix in fig 6 that regression correctly predicted 51% of the churners and 94% of the non-churners, with an overall accuracy of 81%. It is interesting that the linear regression model outperformed all over technologies for predicting non-churners but was not as accurate at predicting the more difficult churn customers.

VII. Comparison of Results

It is apparent that each of the technologies provides varying results. The results for all methods can be viewed in Fig. 7.

	Accuracy % Predicted Churn	Predicted Non-Churn	Overall Accuracy
NN FF/BP Bayesian Architecture	70%	75%	74%
NN FF/BP Standard Architecture	55%	79%	72%
Decision Tree	66%	88%	82%
Regression	51%	94%	81%

Fig. 7 Overall Comparison of Technologies

It can be observed from Fig. 7 that neural networks with a Bayesian framework perform the best for predicting customer churn while linear regression is very accurate for predicting non-churn. Overall the best performing technology was the regression tree and the poorest was the standard neural network.

The most difficult class of customers to predict and the most important, are the churners. This means that the most accurate method for the author's research is the neural network using Bayesian architecture. The table in Fig. 8 compares the most significant variables of all three technologies.

Variables / ***Technology***	Neural Networks	Regression Trees	Linear Regression
No Of Appointments made			●
No. of Repairs			●
Promise Overdue			●
Repair Transfer Count			●
Duration of Repair			●
Complaint Type		●	●
No. of Complaints		●	●
No. of Broken Promises		●	
Cash Conceded		●	
No. Missed Appointments		●	
No. Repair Appointments	●	●	
If an Order Has Been Made	●	●	
Repair Has Been Transferred	●		
No. of Engineers On Site	●		
Resolution Time	●		
Duration of Repair	●		
Length of Custom	●		

Fig. 8 Most Significant Variables

As can be viewed from Fig. 8, four of the variables were used by more than one of the technologies. Complaint type represents information about the sort of complaint which was reported. No. of complaints records how many times the customer has logged a complaint with the company. These two variables were used by both the regression tree and linear regression models. No. repair appointments records how many appointments were taken to complete a repair. The variable if an order has been placed records if a customer has placed an order to repair a fault. These two variables are used by both neural networks and regression trees.

From this information we can assume that from the twenty four original variables, the most significant are complaint type, No. of complaints, No. of repair appointments and if an order has been placed as these are used by multiple technologies.

VIII. Future Research

The next stage of the author's research will involve performing a deeper analysis into the customer data to try to establish new variables that will enhance the predictions of the technologies based on the existing ones. For example, if a customer has made several complaints throughout the year, the author's could use standard deviation to create a new variable that records how often the customer complains in terms of average time intervals. It is possible that a customer who has made four complaints may differ from another customer who has made the same number of complaints if the complaints where made during a short amount of time or spread over the whole year. The author's assume that experiments such as this could offer some interesting effects on the accuracy of the predictions.

With the most suitable technology for predicting customer churn identified, the author's future research directions will focus on creating customer lifetime profiles based on the loyalty index rates defined by the neural network. These profiles will allow the company to fit their customer base into *n* categories and make a long estimation on when a customer is potential going to terminate their service with the company.

An interesting observation seen in Fig. 7 is the fact that the best overall technology for predicting customer churn/non-churn is decision trees; however the neural network with Bayesian architecture was the best technology for predicting churners, while the basic regression model was the best at predicting non-churners. Based on this, the authors would like to investigate a two-wave churn prediction approach where the customer base is first analysed for non-churners using a regression model and then analysed again for churners using the neural network model. Undoubtedly this would lead to certain customers appearing in both the churn and non-churn groups, therefore a third group of 'fuzzy customers' would be created. An analysis and investigation into this idea would allow the authors to make a decision on if these customers should belong to the non-churner group, churner group or remain in their own group as customers requiring none or immediate attention.

Further to the profiling of the customers a real time environment will be developed to constantly monitor the customer's interactions with the company and dynamically shift a customer to a more appropriate profile if needed.

The authors would like to investigate the use of fuzzy clustering methods for analysing the customer base as an extension to the neural network, decision tree and regression experiments that have already been performed. This will be investigated in due course as long as time is permitting.

IX. Conclusion

This research has provided some interesting results and insights into predicting customer churn and the varying technologies available for the purpose of prediction. As mentioned, it is apparent from Fig. 7 that the decision tree is the best overall technique; this is probably because of its rule based architecture. The decision tree consists of a complex rule set and filters customers down the rules (the tree) to either churn or not churn. The process is more on classifying the customers into two groups therefore it can take into account both churn and non-churn. The neural networks and regression have been trained to make calculations to decide if customers are churners. The accuracy depends mainly on the weights for the neural networks and coefficients for the regression. By playing around with these figures it maybe possible to boost the accuracy of the technologies however manual adjustments would be time consuming and difficult.

From the fourteen variables used by linear regression only two of the variables where also present for the classification tree, cash conceded and no. of complaints made. Several assumptions can be made from this comparison. First we can assume that the most important variables from our complaints data are cash conceded and the number of complaints made by a customer and secondly we can determine that just because certain data works with one technology it doesn't necessarily mean it will work best with another.

This work has addressed several research gaps. It was identified from the literature that most researchers focus churn prediction on usage and demographic data. The author's research shows significant accuracy for predicting customer churn using repairs and complaints data, proving that repairs and complaints influence customer's decisions to stay with their service providers.

Earlier research has focused on simplistic classification of customers as churners and non-churners. This research has extended previous research by modelling the churn risk and loyalty index in a continuous fashion. This research has also identified the key variables that affect churn. In addition, it has proposed an accurate and validated model for churn prediction.

References

[1] Au, W., Chan, C. C. and Yao, X. (2003) A Novel Evolutionary Data Mining Algorithm with Applications to Churn Prediction, IEEE transactions on evolutionary computation, 7, 6, 532-545.

[2] Baesens, B., Verstraeten, G., Van Den Poel, D., Egmont-Peterson, M., Van Kenhove, P. and Vanthienen, J. (2004) Bayesian Network Classifiers for Identifying the Slope of the Customer Lifecycle of Long-Life Customers, European Jornal of operational Research, 156, 508-523.

[3] Behara, R. S., Fisher, W. W. and Lemmink, J. (2002) Modelling and Evaluating Service Quality Measurement Using Neural Networks, International journal of operations and production management, 22, 10, 1162-1185.

[4] Bloemer, J., Brijis, T., Vanhoof, K. and Swinnen, G. (2002) Comparing Complete and Partial Classification for Identifying Customers at Risk, International journal of research in marketing, 20, 117-131.

[5] Boone, D. S. and Roehm, M. (2002) Retail Segmentation Using Artificial Neural Networks, International journal of research in marketing, 19, 287-301.

[6] Chan, P. K., Fan, W., Prodromidis, A. L. and Stolfo, S. J. (1999) Distributed Data Mining in Credit Card Fraud Detection, Intelligent Systems and Their Applications, IEEE, 14, 6, 67-74.

[7] Datta, P., Masand, B., Mani, D. R. and Li, B. (2001) Automated Cellular Modeling and Prediction on a Large Scale, Issues on the application of data mining, 14, 485-502.

[8] Dudoit, S. and Van Der Laan, M. J. (2003) Asymptotics of Cross-Validated Risk Estimation in Estimator Selection and Performance Assessment, Statistical Methodology, 2, 131-154.

[9] Fong, J. and Cheung, S. K. (2004) Translating Relational Schema into Xml Schema Definition with Data Semantic Preservation and Xsd Graph, Information and software technology, 47, 437-462.

[10] Ho Ha, S., Min Bae, S. and Chan Park, S. (2002) Customer's Time-Variant Purchase Behavior and Corresponding Marketing Strategies: An Online Retailer's Case, Computers and Industrial Engineering, 43, 801-820.

[11] Hsieh, N. (2004) An Integrated Data Mining and Behavioural Scoring Model for Analysing Bank Customers, Expert systems with applications, 27, 623-633.

[12] Hwang, H., Taesoo, J. and Euiiho, S. (2004) An Ltv Model and Customer Segmentation Based on Customer Value: A Case Study on the Wireless Telecommunications Industry, Expert systems with applications, 26, 181-188.

[13] Kavzoglu, T. and Mather, P. M. (2001) The Role of Feature Selection in Artificial Neural Network Applications, International Journal of Remote Sensing, 23, 15, 2919-2937.

[14] Kim, H. and Yoon, C. (2004) Determinants of Subscriber Churn and Customer Loyalty in the Korean Mobile Telephony Market, Telecommunications Policy, 28, 751-765.

[15] Lee, H. (1999) Semantics of Recursive Relationships in Entity-Relationship Model, Information and software technology, 41, 877-886.

[16] Meyer-Base, A. and Watzel, R. (1998) Transformational Radial Basis Neural Network for Relevant Feature Selection, Pattern Recognition, 19, 1301-1306.

[17] Rosset, S., Neumann, E., Uri, E., Vatnik, U. and Idan, Y. (2002) Customer Lifetime Value Modelind and Its Use for Customer Retention Planning, Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 332-340.

[18] Rygielski, C., Wang, J. and Yen, D. C. (2002) Data Mining Techniques for Customer Relationship Management, Technology in Society, 24, 483-502.

[19] Shin, H. W. and Sohn, S. Y. (2004) Segmentation of Stock Trading Customers According to Potential Value, Expert systems with applications, 27, 27-33.

[20] Van Den Poel, D. and Lariviere, B. (2003) Customer Attrition Analysis for Financial Services Using Proportional Hazard Models, European journal of operational research, 157, 196-217.

[21] Vellido, A., Lisboa, P. J. G. and Meehan, K. (1999) Segmentation of the Online Shopping Market Using Neural Networks, Expert systems with applications, 17, 303-314.

[22] Wei, C. and Chiu, I. (2002) Turning Telecommunications Call Details to Churn Prediction: A Data Mining Approach, Expert systems with applications, 23, 103-112.