



Working with telecommunications

Churning in the telecommunications industry

Churn is the process of customer turnover. In the mobile telecommunications market, a 1995 report by Digital Equipment Corp. estimated the cost of churn at around \$400 per new subscriber. Churn is not restricted to the telecommunications market – it occurs wherever stiff competition provides incentives for customers to switch providers. Industries such as credit card issuers, insurance companies and ISP’s are all subject to increasing levels of churn. ISP’s have hinted at churn rates as great as 50 percent, the mobile telecommunication rate appears to vary around the 30 percent mark. As the market plateaus in these industry sectors (save perhaps for ISP’s) the problem of churn is becoming ever more critical.

Churn analysis using data mining

In order to combat the high cost of churn, increasingly sophisticated techniques can be employed to analyze why customers churn and which customers are most likely to churn in the future. Such information can be utilized by marketing departments to better target recruitment campaigns and by active monitoring of the customer call base to highlight customers who may, by the signature in their usage pattern, be thinking of migrating to another provider.

Data mining techniques bring powerful modeling analytics to the identification of the “Why” and “Most Likely” groups. This paper describes a typical application of such techniques to a mobile telecommunication provider’s customer base.

Customer churn analysis with Clementine

Clementine is a knowledge discovery workbench recognized as an industry leader in the integration of data mining technologies with ease of use. Based on an intuitive visual data flow process of analysis, models and reports can be created to tackle otherwise difficult and time consuming problems. Consider the following data set:

International	Local	Long_Dist	Pay_Meth	Local_Bill_Type	Local_International_Type	Age	Sex	Status	Children	Est_Income	Car_Owner	Churned
7,313	46,328	0	CR	FreezeCall	Standard	37,701	F	M	2	27535,0	Y	Vol
0	6,71	0	CC	FreezeCall	Standard	37,87	F	M	1	36515,0	N	Current
0	58,293	0	CC	FreezeCall	Standard	57,81	M	M	2	78824,1	Y	Current
0	3,840	0	CC	Budget	Intro_Account	58,504	F	S	2	64432,0	N	Vol
0	9,361	1	CC	Budget	Intro_Account	68,762	F	M	2	91366,0	N	Vol
0	19,117	0	CC	FreezeCall	Standard	41,198	M	S	2	67447,0	N	Current
0	26,440	0	M.A.S	FreezeCall	Intro_Account	35,640	M	M	1	42917,0	N	Current
0	30,817	0	CC	FreezeCall	Intro_Account	64,891	M	S	1	6591,87	Y	Current
3,30	23,897	0	CC	Budget	Standard	24,834	M	S	0	67967,1	Y	Current
0	3,327	0	CC	FreezeCall	Standard	37,599	F	M	0	67888,2	Y	Current
2,354	22,828	0	CC	FreezeCall	Intro_Account	68,809	M	M	2	91220,0	N	Vol
0	1,413	0	CC	FreezeCall	Standard	94,9	F	S	0	54425,0	N	Vol
0	37,809	0	CC	FreezeCall	Standard	32,367	M	M	1	47828,0	N	Current
0	8,537	0	CR	Budget	Intro_Account	28,821	F	M	2	28964,4	N	Vol
0	21,582	2	M.A.S	FreezeCall	Standard	32,121	F	M	1	6704,1	Y	Current
0	18,172	0	CC	Budget	Standard	42,86	F	S	2	32371,4	N	Vol

This data shows us a variety of customer information. The last field, Churned, indicates what the current status of each customer is. We will use this information to find which customers are most likely to be voluntary churners.

The data represents a sample of customers with descriptors such as number of minutes on long distance and local calls, financial information such as billing types and payments methods, some demographics such as age, sex, income. The final field “Churned,” indicates whether the customer has ceased usage on their own accord (Vol), or are still

a current paying customer (Current). The InVol customers are those who have been removed by the telecommunications provider for reasons not due to the churn we are interested in, i.e. payment defaults, etc.

The task is to profile those customers most likely to be voluntary churners.

Types of analysis

Decision trees and neural networks are the staple of advanced knowledge discovery workbenches. These techniques can be effectively applied to the churn problem. Neural networks are able to uncover complex patterns in the types of customers and rank the customer base based on a score or likelihood, to churn. Although powerful, neural networks are difficult to interpret and do not readily give up the process by which a customer has been scored.

Decision trees on the other hand build very open and interpretable models that show the analyst the patterns discovered. For example, a decision tree algorithm may discover the following rule.

If international call time is 10 minutes and the long distance bill type is standard, then churn likelihood **high**.

This rule can lead to actionable knowledge, it may be worth calling this customer and suggesting a change of billing type to one more suited to this level of international calling.

Neural networks, decision trees and even clustering algorithms can be helpful in the analysis of customer churn.

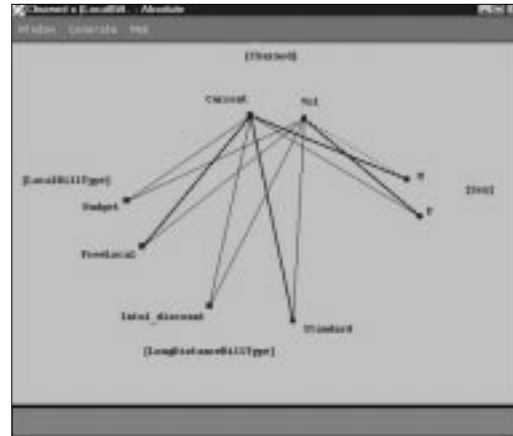
Feature selection

In any data analysis there is a process. Specifically:

- Business question definition (in this case churn analysis – who is churning? And can we develop an action plan to combat that churn?)
- Data gathering – build the data sample. Summarize customer-calling behavior and join with relevant demographics and billing information
- Feature selection – a combination of identifying poorly populated (noisy?) attributes and attributes not deemed relevant to the task at hand. This is done by visually inspecting the data and, together with a domain expert, identifying those attributes most likely to impact the churn decision
- Modeling – build the models
- Analysis and report generation

Clementine™ supports all aspects of the churn analysis process apart from the business question definition. This has to be defined by domain experts, the actual business problem drives the data-mining task. Once data has been gathered from various database sources Clementine can be used to visualize and explore. Histograms, scatter plots and association graphs can be created to help uncover patterns. The following picture shows an association graph (sometimes known as a Web diagram).

The darker the line connecting any two attribute values, the stronger that association. It can be seen that females seem likely to churn (Vol), while standard long distance billing types seem more related to not churning (Current) as does free local bill types. This type of visual exploration of the data is often invaluable in allowing the business user to uncover patterns that the analyst “knew” was in the data – “I always thought our highest churn rate was on that type of billing plan.”



This is an association graph, otherwise known as a Web diagram. This lets you do visual exploration to understand the patterns in your data.

After exploring the data the next stage is to begin modeling. The process of feature selection is important. The problem is that some modeling algorithms cannot cope with large numbers of attributes. In many real data sets a user may have hundreds of attributes and this can lead to poor performance of the modeling algorithm. One technique is to reduce the number of attributes to the core set of highly indicative attributes by using decision trees. Decision trees (such as C5 within Clementine) do not have a problem with large numbers of attributes.

The screen shot to the right shows a Clementine visual data flow used to create the graphics and report included within this paper.



This is a Clementine visual data flow used to create the model.

Once the core set of attributes has been selected (a three step process in Clementine) a neural network (or other scoring algorithm) may be trained using the remaining attributes. Once trained (again an easy process in Clementine – as the setting up and training mechanism are all selected automatically by the system) the customers can be sorted into a ranked list of high probability to churn down to the least likely to churn (the loyal customers).

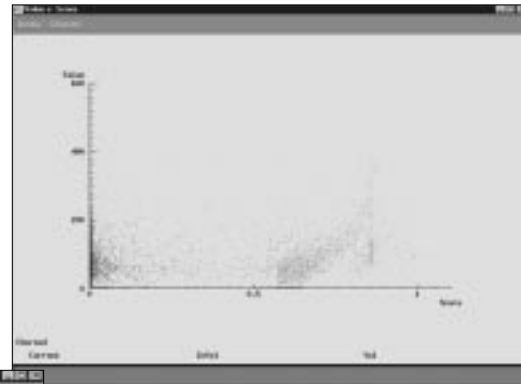
Value and profitability analysis

Even when scoring models have been built, the analysis is by far from over. In some ways it has only just begun. The reason is that there are many ways to use a churn-scoring model – all governed by the business task and problem being tackled. If the goal is to “touch” all customers no matter the cost, then we don’t even need a model. If the goal is

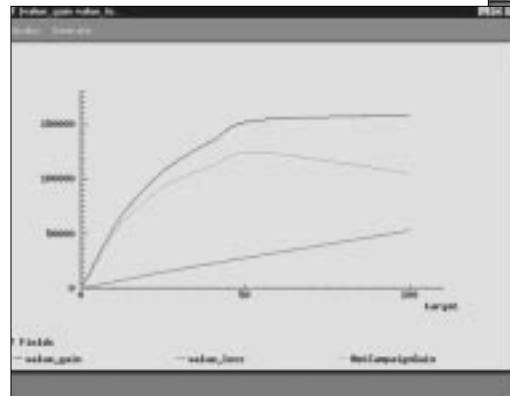
to maximize the return on a marketing budget of a fixed amount, then modeling is critical. Other issues such as customer value begin to play a part. What to do with customers, who are likely to churn, but do not in fact use their mobile phone? If the telecommunications provider is not making money from this customer, should they not just let this customer go? This problem is perhaps even more apparent with credit card issuers.

The chart to the right shows a scatter plot of likelihood to churn versus value.

In Clementine, reports can be created that indicate where to draw the line in customer targeting campaigns, taking into account propensity to churn and customer value (profitability). The following chart indicates a maximum net return in a targeting campaign.

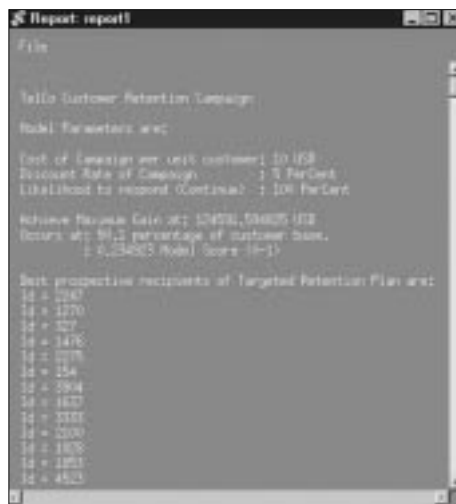


This chart show you a scatterplot of likelihood to churn versus value.



Here, we drew a line to indicate a maximum net return in a targeting campaign.

The middle line shows the net campaign gain based on parameters such as the cost of contacting each customer and the discount rate offered to entice that customer to stay. Other factors relevant to the telecommunications 's particular business problem can easily be added into the Clementine analysis.



Finally, Clementine can generate a list of customer ID's who are the best prospects for a future promotional campaign.

The final report may be a list of customer ID's and a predicted maximum campaign gain, such as the following example. As can be seen, a gain of \$128,226 US dollars is made by targeting (touching) 50.1 percent of the customer base. (Note only a partial list of "hot" ID's is actually shown).

Summary

A brief overview of the use of the Clementine knowledge discovery system in the analysis of customer churn has been presented. For more information please contact SPSS Inc. at www.spss.com.

About SPSS

SPSS Inc. is a leader in the large markets for analytic software: business intelligence (data mining and market research), quality improvement and scientific research. SPSS products and services transform organizations by helping users leverage information to grow revenues and improve processes. More than two million people use SPSS, the world's best-selling tool for desktop analysis, to create and distribute information for better decision making.

Celebrating its 30th anniversary in 1998, SPSS is based in Chicago with more than 40 offices, distributors and partners worldwide. Products run on leading computer platforms, and many are translated into 10 local languages. In 1997, the company employed nearly 800 people worldwide and generated net revenues of approximately \$110 million.

Contacting SPSS

To place an order or to get more information, call your nearest SPSS office or visit our World Wide Web site at www.spss.com

SPSS Inc.	+1.312.651.3000	SPSS Hong Kong	+852.2.811.9662
	Toll-free: +1.800.543.2185	SPSS Ireland	+353.1.496.9007
SPSS Argentina	+541.814.5030	SPSS Israel	+972.9.9526700
SPSS Asia Pacific	+65.245.9110	SPSS Italia	+39.51.252573
SPSS Australasia	+61.2.9954.5660	SPSS Japan	+81.3.5466.5511
	Toll-free: +1800.024.836	SPSS Kenya	+254.2.577.262
SPSS Belgium	+32.162.389.82	SPSS Korea	+82.2.3446.7651
SPSS Benelux	+31.183.636711	SPSS Latin America	+1.312.651.3226
SPSS Brasil	+011.5505.3644	SPSS Malaysia	+60.3.704.5877
SPSS Central and Eastern Europe	+44.(0)1483.719200	SPSS Mexico	+52.5.682.87.68
SPSS Czech Republic	+420.2.24813839	SPSS Middle East & South Asia	+91.80.545.0582
SPSS Denmark	+45.45412000	SPSS Polska	+48.12.6369680
SPSS East Mediterranean & Africa	+972.9.9526701	SPSS Russia	+7.095.125.0069
SPSS Federal Systems	+1.703.527.6777	SPSS Scandinavia	+46.8.506.105.50
SPSS Finland	+358.9.524.801	SPSS Schweiz	+41.1.266.90.30
SPSS France	+33.1.5535.2700	SPSS Singapore	+65.533.3190
SPSS Germany	+49.89.4890740	SPSS Taiwan	+886.2.25771100
SPSS Hellas	+30.1.7251925	SPSS UK	+44.1483.719200
SPSS Hispanoportuguesa	+34.91.447.37.00		