Analyzing Customer Churn in the Software as a Service (SaaS) Industry

Ben Frank, Radford University
Jeff Pittges, Radford University

Abstract

Predicting customer churn is a classic data mining problem. Telecommunications providers have a long history of analyzing customer usage patterns to predict churn. Many other industries, such as banking, routinely analyze customer behavior to predict customer satisfaction and renewal rates. The Software as a Service (SaaS) model enables software vendors to collect customer usage data that is not available to traditional software vendors. The SaaS market, and cloud computing in general, is growing rapidly yet to our knowledge little work has been done to apply existing methods to analyze customer churn in the SaaS industry. This paper describes a study conducted on behalf of a leading performance and talent management company. Our work uses churn analysis in telecommunications as a baseline to study the SaaS industry. Although there are many similarities between the two industries our study identifies several important differences. The paper presents four experiments and outlines opportunities for future research.

Keywords: Churn analysis, churn prediction, cloud computing, customer attrition, software as a service.

1 Introduction

Software as a Service (SaaS) is a deployment model whereby customers subscribe to a service rather than purchasing a license to own a software product. Subscriptions are sold for a period of time (e.g., monthly or yearly subscription) similar to cell phone service and utilities. The SaaS model offers many advantages to customers. The low cost of a subscription makes high-end products available to companies that could not afford to purchase the product. System administration is managed by the vendor which eliminates expenses and headcount for the customer and reduces time and expenses for training and support. The recurring revenue stream of subscription based services is attractive to vendors, but the vendor's dependence on subscription renewals makes the business far more sensitive to customer satisfaction.

When a customer purchases a traditional software license the product is installed on the customer's servers and the vendor has no visibility into how the customer is using the product. SaaS vendors host their applications by running the software on servers managed by the vendor. This enables vendors to collect valuable data on how customers are using their product including who is using the product, when, for how long, and how often. Many industries, such as telecommunications and banking, already rely on usage patterns to predict customer churn and customer satisfaction. SaaS vendors have the ability to predict similar characteristics of their customers.

Customer churn in service industries has been studied extensively. However, to our knowledge, Software as a Service (SaaS) has not been a specific focus. Although the SaaS industry shares many commonalities with telecommunications and other service industries, there are some key differences worth exploring especially given the rapid growth of the SaaS market and the increasing availability and reliability of cloud computing.

This paper describes a one-semester student project to study customer churn in the SaaS industry. The study was conducted on behalf of a leading performance and talent management company. The paper draws comparisons between the SaaS and telecommunications industries, describes four experiments that apply existing methods to predict churn, and presents opportunities for future research to better understand customer churn in the SaaS industry.

2 Related Work

Customer churn is a classic data mining problem and many approaches have been developed for the telecommunications industry. The following work provides a foundation for our current and future research.

Euler [3] developed a decision tree to identify types of telecommunications customers most likely to churn. Euler utilized the data preprocessing capabilities of the MiningMart KDD system to derive predictive features that were not present in the original data. Derived values were an important component of our final and most successful model. Euler's model also incorporates temporal aspects of customer behavior and we plan to use a similar approach to measure the velocity of several features as we continue our research.

Coussement and Van den Poel used support vector machines to improve the performance of predicting churn for a newspaper subscription service [1]. The results of this work show that interactions between the clients and the provider are an important predictor of churn. Coussement and Van den Poel continued their study of client/provider interactions by adding emotions from client emails to their model [2]. Hadden et al [4] identified predictive features of customer complaints and found that decision trees outperform neural networks and regression in terms of overall accuracy.

3 Project Goals and Hypotheses

The project was conceived with two primary stakeholders, a student and the vendor. The student had completed a course on data warehousing, data mining, and reporting [5] and wanted to learn more about data mining. The vendor wanted to predict customer satisfaction and renewal rates.

The study was designed as an exploratory project to understand the vendor's data and to evaluate existing methods for analyzing customer churn. The primary objective was to identify the most important features of usage data that predict customer satisfaction.

The project was guided by the following hypotheses:

1. Deployment (percentage of employees with accounts) is a significant measure of customer satisfaction.
2. Deployment rate (how quickly user accounts are created) will help account managers measure satisfaction and identify stalled engagements.
3. Adoption rate (deployment and usage) is a significant measure of customer satisfaction.
4. Usage across levels of the organization (e.g., executives, middle managers, and independent contributors) should indicate stronger adoption and overall satisfaction.

This study was conducted in collaboration with a leading provider of performance and talent management services. The vendor offers several modules from which a customer may choose. To properly use these services a customer must purchase a license for every employee to access the applications hosted by the vendor. Customers typically purchase all licenses (one per employee) when the contract is signed, but each module is deployed in phases, one department at a time. This practice allows the vendor to monitor product adoption by observing two variables for each customer: the total number of licenses purchased and the number of user accounts that the customer has created for their employees.

Our first hypothesis is that the rate of deployment, how quickly the customer creates accounts for their employees, should provide a strong indication of the customer's satisfaction level. Furthermore, a decrease in deployment rate may indicate that the customer is becoming dissatisfied with the product. Establishing a baseline for deployment rate with key milestones based on the percentage of users would enable account managers to measure satisfaction, identify stalled accounts, and take action proactively.

The last two hypotheses pertain to the overall satisfaction of the customer. It is not uncommon for executives to purchase a productivity tool only to find that the employees are not using the product. The adoption rate measures how much the product is being used. However, usage must be based on the percentage of users who have accounts (i.e., deployment). The final hypothesis takes into consideration the diversity of the user population by grouping users by rank. High usage across all levels of an organization should indicate high overall satisfaction. Furthermore, identifying groups with low usage enables the vendor and the customer to address issues with a particular group.

4 Comparing Data from SaaS and Telecommunications Providers

The vendor provided aggregate, anonymized usage data collected over a multi-year span. The data includes the 8 attributes described in Table 1. These 8 attributes are collected daily for individual users (employees) from over 600 customers. The percentage of valid users is calculated by dividing the number of valid users by the total number of licenses (max_seats). The data is loosely described in accordance with our confidentiality agreement with the vendor.

Attributes	**Description**
all_users	The total number of times a module was accessed
cust_date	Date the customer was acquired
dist_users	The number of distinct users to access a module
logins	Number of times the customer logged in
max_seats	Total number of licenses purchased by the customer
mkt_segment	Customer's market segment
module	Name of the module used
valid_users	Total number of valid users for a customer

Table 1: Usage data attributes.

There are many similarities between the telecommunications industry and the SaaS industry. As shown in Table 2, both industries are capable of tracking usage with respect to the number of times their product is used and the length of time their products are used. Telecommunications providers record the details of each call while SaaS vendors record the details of each session. A session begins when a user logs into the system (login) and the session ends when the customer logs off or when the session times out due to inactivity. Therefore, while most callers are active during a call, except for hold time, software users may not be active during an entire session. When a session times out the timeout period should be subtracted from the length of the session to more accurately measure the time the user was actively using the system.

Telecom	**SaaS**
Number of calls	Number of logins (sessions)
Length of call	Length of session
Call targets	Modules used

Table 2: Comparison of Telecom data with SaaS data.

Telecommunications providers often track who their customers are calling (e.g., ISPs, toll free numbers). SaaS vendors typically offer a suite of products and they track which modules are being used. Telecommunications providers offer a wide variety of applications whereas SaaS vendors typically offer a related suite of products. Hence, SaaS vendors typically have more focused offerings which has advantages for identifying product adoption.

Market segments apply to both industries. However, much of the research on customer churn in telecommunications has studied business-to-consumer (B2C) relationships whereas the SaaS vendor in our study only sells to businesses (B2B). Therefore, while the customer attributes may be similar, our study uses data aggregated over all of a customer's employees. Furthermore, we are able to group a customer's employees by rank and other properties which may lead to important insights.

5 Experiments

Four experiments were conducted to understand the vendor's data. Each experiment built models with the Weka data mining suite [6]. The models were built with data from over 600 customers known to be satisfied or unsatisfied. Approximately 60 percent of the customers in the training set were preclassified as satisfied.

5.1 Experiment One: Clustering

The first experiment attempted to further classify customers beyond satisfied and unsatisfied using the K-means algorithm to cluster customers into 2, 3, and 5 groups. All three iterations produced a single cluster that contained over 70 percent of the customers.

We chose to cluster the customers into two groups expecting the K-means algorithm to separate the customers into satisfied and unsatisfied, but this was not the case. Unsatisfied customers were almost evenly distributed between the two clusters. We tried three clusters expecting a group form between the original two clusters but the single large cluster was mostly unchanged. The model with five clusters also failed to break up the large group and none of the three models produced a cluster that could be interpreted in any meaningful way.

5.2 Experiment Two: oneR

The remaining experiments built decisions trees to identify the most relevant attributes for predicting customer satisfaction. A third of the customers in the training set were randomly chosen to build each tree and the remaining two-thirds were used to test the models.

Experiment 3	Experiment 4
cust_date	percent_valid_users
max_seats	dist_users
dist_users	percent_valid_users
valid_users	all_users
max_seats	module
all_users	logins
logins	dist_users
max_seats	all_users
	percent_valid_users

Figure 1: Decision tree attributes

In the second experiment, the oneR algorithm identified the number of valid users as the most important attribute for predicting customer satisfaction. This is consistent with our expectations that product adoption and retention depend on customer deployment. However, we also hoped to find a distinct number of users, a threshold, that would partition the majority of customers into satisfied and unsatisfied. The vendor could use this threshold as a target for all of their customers. Unfortunately, multiple sections of satisfied and unsatisfied customers were scattered throughout the range of valid users.

5.3 Experiment Three: J48

For the third and fourth experiments we created J48 decision trees. The first tree was built with all eight attributes. The final pruned tree had 8 levels with 23 leaves and a size of 45. The attributes at each level of the tree from the root down are shown under Experiment 3 in Figure 1. Neither of the top two attributes, customer acquisition date (cust_date) and the maximum number of users (max_seats) are actionable, as the data is historical and the vendor does not want to focus solely on large accounts.

5.4 Experiment Four: J48

The first J48 decision tree identified acquisition date and maximum seats as the most important predictors of customer satisfaction but neither attribute is actionable. In the fourth experiment we derived more meaningful values based on the acquisition date and the maximum number of seats purchased by the customer.

Customer acquisition date is a single, fixed point in time that does not account for the length of time the customer uses the product. The customer acquisition date was transformed to the number of months the customer has licensed one or more modules. By measuring the length of the engagement we hoped to identify a significant milestone for product adoption. This would produce a valuable target for the vendor.

The maximum number of seats is also a fixed number biased towards larger customers. By combining the maximum number of seats with the number of valid users we derived the percentage of valid users which better reflects adoption across the entire company regardless of size.

The modified data set produced a tree with 9 levels, 39 leaves, and a tree size of 69. The attributes at each level of the tree from the root down are shown under Experiment 4 in Figure 1. The second J48 decision tree correctly classified 96 percent of the customers in the test data set. No conclusive threshold emerged for the length of engagement, but the results indicate that a series of milestones over time may produce an adoption rate baseline consistent with our third hypothesis.

The top three attributes in the second J48 tree are the percentage of valid users, the number of distinct users, and the total number of times the system has been accessed. The percentage of valid users measures the degree to which management has adopted the product. The number of employees using the system (dist_users) and the number of times the system has been used (all_users) measure the degree to which the employees have adopted the product. Finding these three attributes at the top of the tree is consistent with our first and third hypotheses.

Everyone involved in the study expected market segment to have a high impact on customer satisfaction. Our hypothesis was that software applications would appeal to customers in some markets more than others. The fact that market segment was pruned from the final tree is likely an indication that software has become an integral part of doing business.

6 Future Research

There are many opportunities to further analyze usage data from software vendors employing the SaaS model. We plan to build on our fourth experiment by deriving additional variables and incorporating time, similar to Euler [3]. For example, computing the percentage of valid users who have used the system (dist_users / valid_users) should provide a more precise measure of employee adoption. As we create additional derived values it may be worthwhile to experiment further with clustering. We also want to build a oneR model using the derived data from the fourth experiment as we are more likely to find a threshold using the percentage of valid users in place of the number of valid users.

Time is an important factor for measuring rates of deployment and adoption. Tracking the number of valid users who use the system over time, both fixed (monthly) and sliding (last 30 days) intervals, will enable us to identify trends. Developing a temporal model to identify adoption milestones, such as percentage of valid users and percentage of distinct logins, and the time periods that indicate satisfaction or dissatisfaction, would enable the vendor to manage product deployment more proactively.

As discussed in Section 2, Coussement and Van Den Poel found client/provider interactions to be an important predictor of customer satisfaction [1, 2]. Incorporating complaints data into our models should increase reliability in accordance with the findings of Hadden et al [4]. As suggested in [4], time is also important for measuring the frequency of complaints.

Segmenting employees also appears promising. As described in our fourth hypothesis, tracking usage by rank should provide further insight into overall satisfaction. What is the ideal mix of usage across the levels of an organization? Do renewal rates depend more on management usage or employee usage? Customers should also be segmented into additional classes of satisfaction. The ability to estimate a customer satisfaction score would be even more valuable as this would allow vendors to determine a customer's direction and velocity along the satisfaction scale. Not only would this help the vendor identify customers at risk of churning, it would also help vendors identify factors that lead to dissatisfaction.

7 Conclusion

Both stakeholders, the vendor and the student, benefited from the project. The study provided support for our first three hypotheses: the rate of deployment and adoption are significant predictors of customer satisfaction. Time constraints prevented us from testing our fourth hypothesis pertaining to adoption across levels of an organization, but we intended to address employee segments in our future research.

Our experiments produced several surprises. The results of clustering customers into 2, 3, and 5 groups were particularly disappointing as none of the clusters could be interpreted in a meaningful way. Clustering with derived values may produce more interesting results. The oneR model confirmed our hypothesis that the number of valid users is an important predictor of satisfaction, but the oneR model failed to produce a single threshold for the vendor to target. We are hopeful that the percentage of valid users may produce a threshold.

Although there are many similarities between the telecommunications and SaaS industries, our study identified several important differences such as telecommunication's focus on consumers versus the business-to-business relationships typical in the SaaS market. Further study is likely to identify additional differences that may be exploited to develop better models for predicting customer churn in the SaaS industry.

Through the process of analyzing the problem, developing a plan of attack, and deciphering the results, the student gained valuable experience with data mining and learned the importance of data preparation and the challenges of working with real data. We will continue to involve students in this research to create excitement among the next generation of data miners.

Acknowledgements

The authors wish to thank Jeff Ulrich for his help and support throughout this project and Bill Eberle for reviewing the paper and providing several helpful comments and suggestions.

References

[1] Kristof Coussement and Dirk Van den Poel. Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications*, 34:313 – 327, 2008.

[2] Kristof Coussement and Dirk Van den Poel. Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating multiple classifiers. *Expert Systems with Applications*, 36:6127 – 6134, April 2009.

[3] Timm Euler. Churn prediction in telecommunications using miningmart. *Proceedings of the Workshop on Data Mining and Business (DMBiz) at the 9th European Conference on Principles and Practice in Knowledge Discovery in Databases (PKDD)*, 2005.

[4] John Hadden, Ashutosh Tiwari, Rajkumar Roy, and Dymtr Ruta. Churn prediction using complaints data. *Proceedings of World Academy of Science, Engineering and Technology*, 13, May 2006.

[5] Jeff Pittges. An undergraduate course in data warehousing. *Proceedings of the 44th Annual Meeting of the Southeast INFORMS*, October 2008.

[6] Weka data mining software. `http://www.cs.waikato.ac.nz/ml/weka/`.

[7] Yaya Xie, Xiu Li, E. W. T. Ngai, and Weiyun Ying. Customer churn prediction using improved balanced random forests. *Expert Systems with Applications*, 36:5445 – 5449, April 2009.

[8] Yu Zhao, Bing Li, Xiu Li, Wenhuang Liu, and Shouju Ren. Customer churn prediction using improved one-class support vector machines. In *Advanced Data Mining and Applications*, pages 301 – 306. Springer Berlin / Heidelberg, 2005.