Using the Future to Segment Your Customers

Robert Do

Introduction

Managers often perform segmentation studies using popular clustering techniques such as K-means clustering. Oftentimes the only data they can cluster upon is the observed behavior of users from the past up until the present. Ideally, a manager would like to be able to segment based on users' future behavior, as decisions are to be made for the future. Since managers cannot look into the future, their best guess at the future is that users will behave as they did in the past. For example, a user who buys four times in the past 10 weeks will similarly buy four times in the following 10 weeks. I present a method that allows for segmentation based on users' future behavior using the Pareto/NBD to provide accurate forecasts of such behavior.

The Data

The data come from an online service that allows event holders to charge for tickets online. Each record represents the number of days in each time period a user logged on to create an event. This will simply be referred to as activity from here on. The sample consists of the activity over 69 weeks of all users who joined in January 2009. The first ten weeks are set aside as the model calibration period. The remaining 59 weeks are used as the holdout data. Outliers were removed to avoid the creation of segments of very small size. Each record has four pieces of information: the frequency of activity in the calibration period, the week of the most recent activity, the length the user has been observed, and the frequency of activity in the holdout period. Below is a sample of the data.

ID	p1x	tx	T	p2x
493836	0	0	10	0
493856	1	0.714286	10	1
493871	0	0	10	0
493878	0	0	10	0
493880	0	0	10	0

p1x =	Activity in the calibration period
tx =	Week of most recent usage
T =	Length of observation
p2x =	Activity in the holdout period

K-Means Clustering Analysis For Segmentation

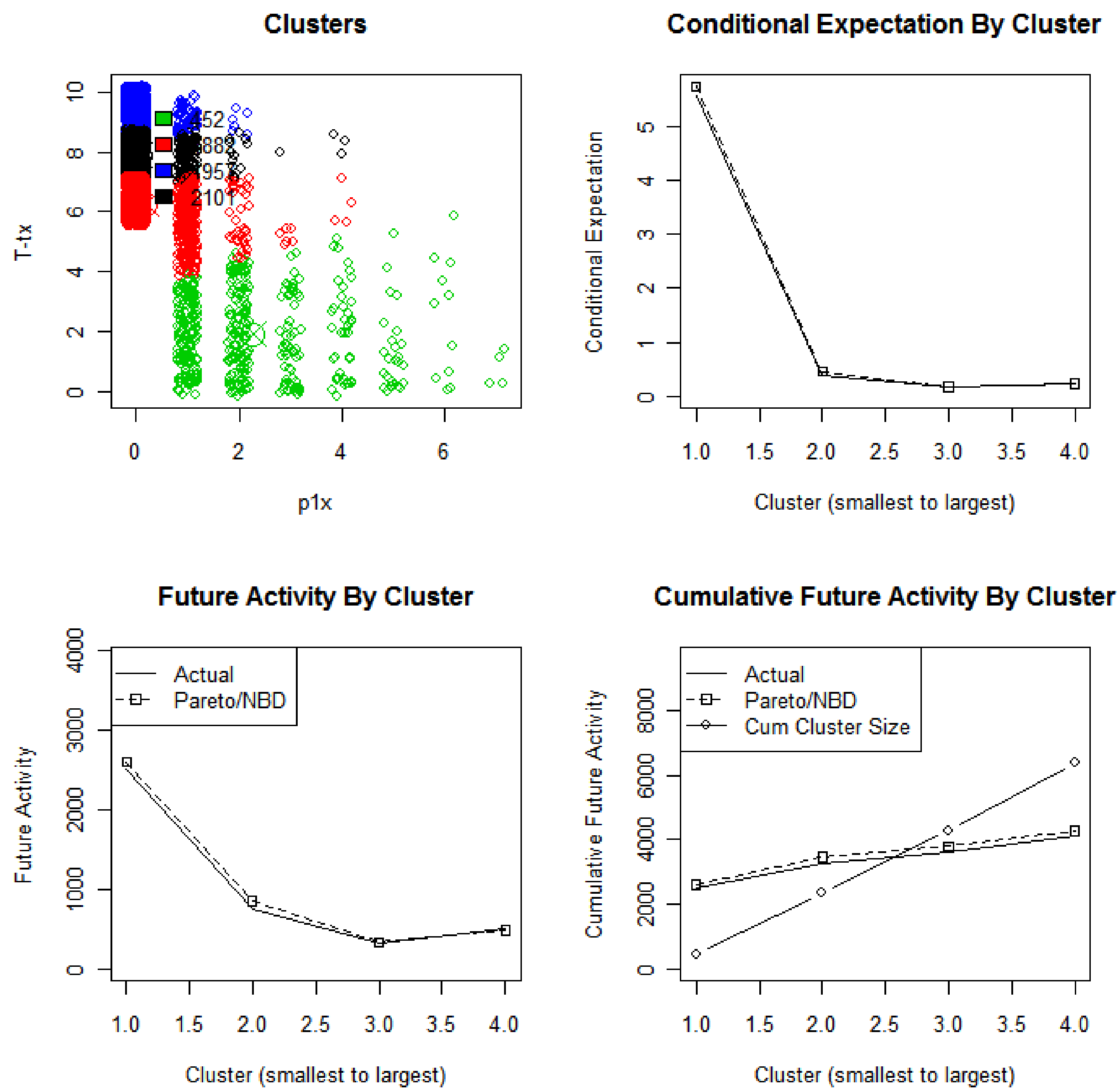
K-Means clustering analysis is a popular data mining technique commonly used to partition observations into k number of clusters or segments. Usually the clusters are formed based on observable characteristics of the users such as purchase frequency, purchase recency, total amount purchased, and demographic information. This results in clusters of people of who are most similar to each other in their respective cluster and most dissimilar to others outside their cluster. However, since these groups were created based on past data, the groupings may seem rather obvious and may not hold true in the future.

This would be analogous to assuming a baseball player's batting average in a single game is the same as his true long run batting average. He may have been lucky in that game and his single game average was higher than usual or he may have been unlucky and it was lower than usual. When team managers pick players, they do not rely on single game batting averages, but on long run averages. Managers should do the same when grouping users. They should not rely on the first 10 week usage frequency, but rather on the expected usage in the future since this analysis is to be used in making decisions for the future.

Many managers cannot do this because they cannot see into the future. However, predicting the future can be done if the holdout data can be accurately forecasted. The proposed Pareto/NBD model can accurately forecast future activity and can provide conditional expectations for groups of users. Thus, rather than cluster based on the past frequency data, we cluster based on the conditional expectations. This method takes into consideration the unobservable latent characteristics that are more indicative of true underlying user behavior.

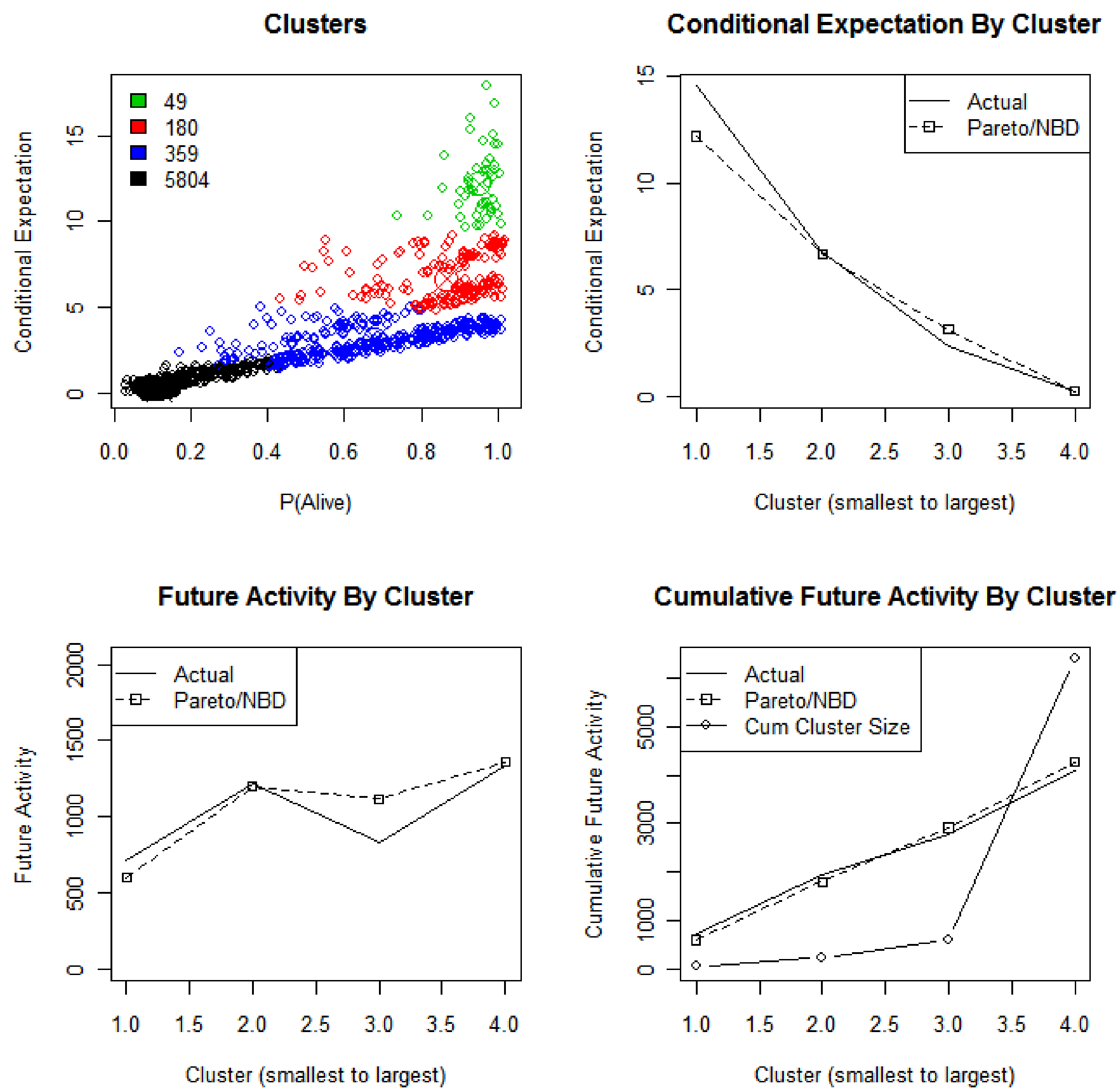
A comparison of clustering based on observable characteristics of frequency and time since last activity and clustering based on conditional expectation is presented below.

Clustering based on observable characteristics (p1x, T-tx)



Cluster	1 (active)	2 (triers)	3 (triers)	4 (triers)
Size	452	1882	1962	2096
Conditional expectation (expected usage over next 59 weeks)	5.75	0.45	0.23	0.17
Total expected usage	2598	855	457	361
Time between each usage	2.4 months	30.0 months	58.4 months	79.1 months

Clustering based on conditional expectation



Cluster	1 (high)	2 (medium)	3 (low)	4 (triers)
Size	49	180	359	5804
Conditional expectation (expected usage over next 59 weeks)	12.19	6.66	3.11	0.23
Total expected usage	597	1198	1118	1357
Time between each usage	1.1 months	2.0 months	4.4 months	58.2 months

This comparison demonstrates how the latter method has more discriminatory power in identifying distinct segments. The former method makes pretty accurate predictions, but when there are more than

2 segments, the other segments are more or less the same. For example, segments 2, 3, and 4 more or less fall into the same bucket as less important customers due to their similar low usage. There is pretty much only 2 segments, active users and super low activity users. The conditional expectation method can identify more distinct segments. Ideally, we would like to be able to break “active users” down into more descriptive segments such as high, medium, and low.

Using conditional expectation allows segmentation of users based on their expected future behavior rather than merely their past behavior. When the model can accurately predict conditional expectations, this method of clustering is extremely powerful. It is as if you are segmenting based on the holdout data. This segmentation produces managerially meaningful segments that describe implications of future activity such as users who WILL use the product every 1.1 months, users who WILL use it every 2 months, users who WILL use it every 4.4 months, and users who WILL use it every 58.2 months (essentially almost never). These segments can be appropriately named high, medium, low, and triers. Although each member of the last group has very low expected future usage, the group should not necessarily be ignored because there are so many members in the group. The large number of people (5804 or 91% of users) combined with a low future expected usage still results in 1357 usages (33% of total activity).

Solving the 80/20 Problem

A common managerial question is the 80/20 problem. “Which 20% of my customers make up 80% of my sales?” Usually this analysis is done on past sales data when the answer is to be used for decisions made for the future. The proposed method provides the answer to the more appropriate question, “Which 20% of my customers WILL make up 80% of my sales?” This data set exhibits a 10/70 behavior where 10.3% of users represent 70.7% of activity. This can be found by creating 10 clusters based on conditional expectation and filtering on the clusters that have a conditional expectation above a threshold (in this case 1). The threshold is chosen as one that results in a customer with the least activity worth serving. A threshold of 1 corresponds to a user who uses the product once every 13.6 months or about 1 year. A threshold of .5 corresponds to a user who uses the product once every two years and who is probably not worth the cost of serving.

How Are We Predicting The Future?

The key to this clustering method is the ability to predict the future. Below I briefly outline traditional approaches and then describe the more appropriate approach used in this model.

Traditional and wrong approaches

In forecasting sales into the future, some managers make a “best guess” and may simply extend the cumulative sales plot by using their “intuition”, which often means simply taking a pencil and ruler and tracing a line through the data. If they get fancy, they may even bend the ruler. More numbers focused managers may apply the sales growth rate pattern of past weeks into future weeks. More quantitatively sophisticated managers may use statistical methods such as regression. Managers also often try to predict churn for a non-contractual service. It is impossible to tell whether a user has churned in a non-contractual setting because inactivity for a period does not necessarily mean they have left. Sometimes a proxy such as monthly active users is used, but this incorrect as we saw that there are clusters of users who use the product once every 2 months and 4.4 months.

In predicting the future purchasing of a group of customers, managers may assume that a customer who buys 3 times in the past 10 weeks will buy 3 times in the next 10 weeks which is equivalent to the baseball example of assuming a player's batting average in one game will be the same for the next game. Fancier methods may include the use of regression, which may have great in-sample fit, but poor out of sample forecasts.

Understanding the customer usage behavior

The best way to predict future behavior of customers is to truly understand their underlying purchasing behavior. The Pareto/NBD model describes customers of "buying until they die". Die is simply a metaphor for quitting, leaving, etc. "Buy until you die" can be broken down into two behaviors: the buying behavior and the dying behavior. The Pareto/NBD was developed by Schmittlein, Morrison, and Colombo in their paper "Counting Your Customers: Who Are They And What Will They Do Next?" (Schmittlein, Morrison, Colombo 1987).

The Buying Behavior

Each user makes actions according to a Poisson process with rate lambda. The activity rate lambda is different for each user. The activity rate lambda is distributed according to a gamma distribution.

The Dying Behavior

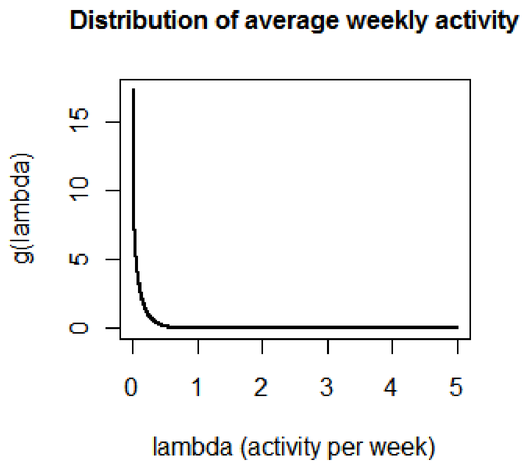
Each user remains alive for a lifetime which has an exponentially distributed duration with death rate mu. The user death rates mu are distributed according to a different gamma distribution.

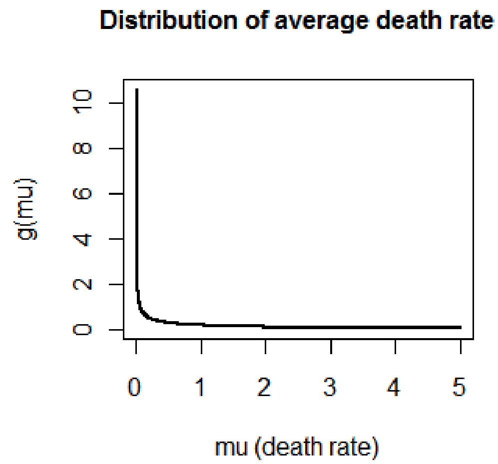
Results of the Model

r	**alpha**	**s**	**beta**	**LL**
0.7730	6.6846	0.4311	0.1391	6213.715

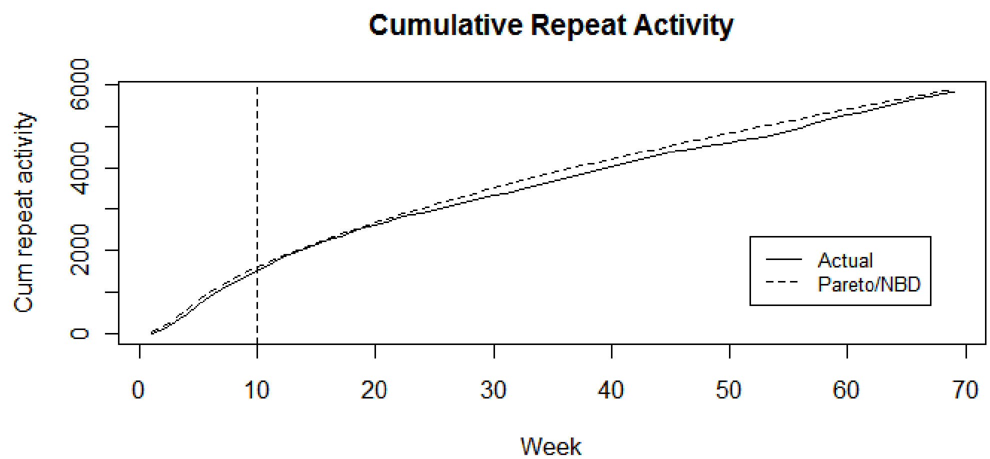
r/alpha = 0.116 activity per week when users are active. The r-value being less than 1 indicates there is heterogeneity among the average activity per week. The plot below demonstrates how much of the weight is on low activity rates.

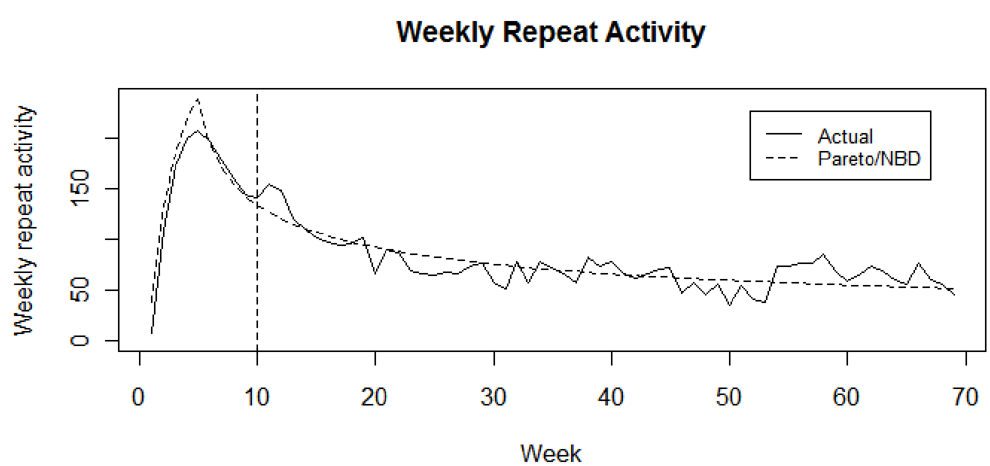
s/beta = 3.099 average death rate. Average lifetime of 1/3.099 = 0.323 weeks. Half the customers will become inactive after $(2^{1/s} – 1)$Beta = 0.555 weeks. The small s-value indicates there is a lot of heterogeneity among the death rates.



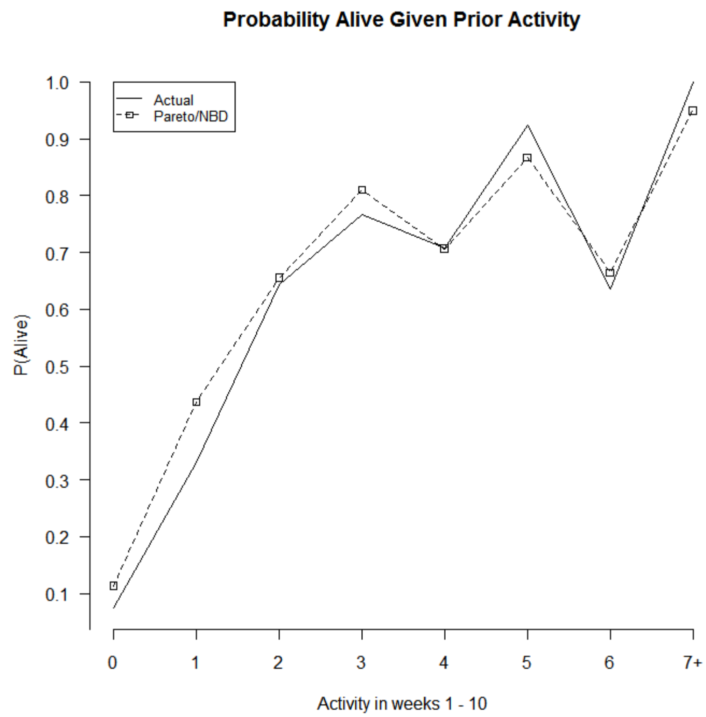


Forecasting Ability





The model forecasts the next 59 weeks extremely well. The out of sample mean absolute percentage error is 3.76% with the week 69 sales forecast off by 1.10%.

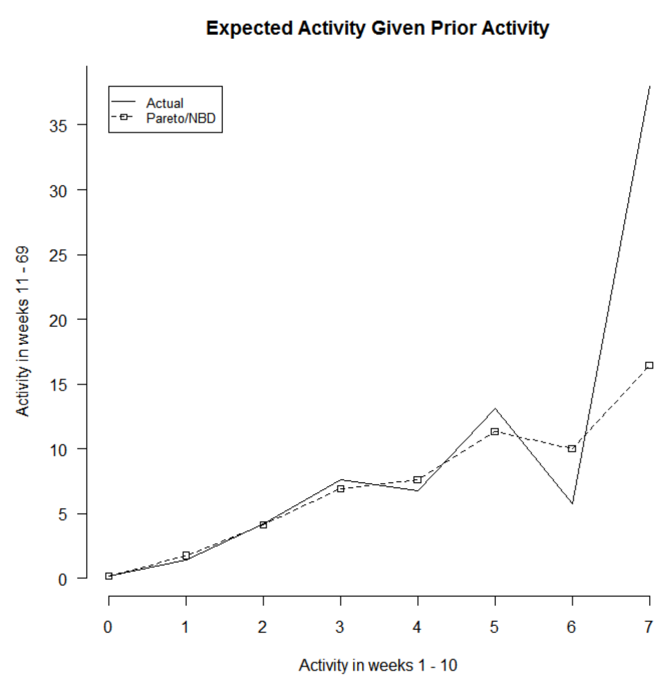


The model is also able to make a best guess at whether a user is still alive at week 10. The actual "aliveness" is impossible to verify because there's no way of knowing if a user is dead or just taking a hiatus from activity. The proxy used to calculate actual P(Alive) is the proportion of users who did any activity at all in the holdout period. If they did something, then they are obviously alive. If they did not do anything in the 59 holdout, we assume they are dead, but they might just be taking a nap!

The model predicts future usage extremely well, especially when there is sufficient number of observations in each group. The prediction for users with 7 usages in weeks 1 – 10 is quite off because there are only four observations in that group, each with widely different tx and T information.

Because the model is able to predict the holdout so well, the conditional expectations can be trusted to be used in clustering.

The full technical mechanics of the model can be found in Schmittlein, Morrison, Colombo (1987). The code for implementing the model was adopted from Faderk, Hardie, Lee (2005).



References

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005), "A Note on Implementing the Pareto/NBD Model in Matlab." <http://brucehardie.com/notes/008/>

Schmittlein, David C. , Donald G. Morrison, and Richard Colombo (1987), "Counting Your Customers: Who They Are and What Will They Do Next?" *Management Science*, **33** (January), 1 – 24.