Cloud of Science. 2014. Т. 1. № 4
http://cloudofscience.ru
ISSN 2409-031X

Моделирование поведения клиентов на примере обслуживания заказов такси

А. Г. Озеров, Р. А. Соловьев

*Компания Wheely (*http://wheely.com*)*
117105, Москва, Варшавское шоссе, 1, стр. 1
e-mail: ao@wheely.com, roman@wheely.com

Аннотация. В статье рассматривается процесс моделирования поведения клиентов в компаниях с непрерывными транзакциями на бесконтрактной основе. Информационной базой для исследования выступают обезличенные клиентские данные сервиса личных водителей службы такси. Рассмотрено регулирование клиентского потока в рамках модели Pareto/NBD. Модель Pareto/NBD позволяет оценить — вероятность того, что пользователь стал неактивным; количество заказов, которые принесет случайный пользователь; количество заказов, которое принесет определенный класс клиентов. В заключении обоснована практическая значимость моделирования для повышения качества и эффективности обслуживания клиентов.

Ключевые слова: Pareto/NBD, моделирование поведения клиентов.

1. Введение

Бурное развитие интернет-технологий в XXI веке ознаменовало появление большого количества сервисов, которые оказывают свои услуги непосредственно или косвенно через интернет. Это не только принципиально изменило схему работы, но и значительно расширило возможности клиентской аналитики. Если раньше доступ к большой базе данных и вычислительным мощностям был только у международных корпораций, то теперь это стало достоянием большинства. Как результат, сегодня можно наблюдать возросший интерес компаний к современным методам анализа данных, позволяющим повысить эффективность и качество своего обслуживания[1].

В статье исследуются особенности моделирования поведения клиентов в компаниях с непрерывными транзакциями и бесконтрактной формой отношений с клиентами. Примером такой компании выступает сервис личных водителей Wheely, обезличенные клиентские данные которого использованы в качестве информационной базы для исследования. Решается задача регулирования клиентского потока в компании Wheely, при условии, что априори неизвестно, отказался ли клиент от сервиса. Решение найдено на основе использования модели Pareto/NBD.

[1]http://www.ibm.com/smarterplanet/global/files/se__sv_se__intelligence__Analytics_-_The_real-world_use_of_big_data.pdf

Таким образом, исследование будет состоять из трех основных частей. В первой части будет произведена проверка модели Pareto/NBD на сгенерированных данных, что позволит нам убедиться в ее общей работоспособности. Вторая часть будет посвящена апробации модели на реальных клиентских данных компании Wheely. В третьей части будут изложены основные результаты исследования и обоснована их практическая значимость для повышения эффективности бизнес-процессов.

Первые исследования, связанные с изучением жизненного цикла потребителя, появились в 1980-х годах. В те годы ключевое положение занимали три модели: NBD (Negative Binomial Distribution), LSD (Logarithmic Series Distribution), NBD/Dirichlet. Они быстро утратили свои позиции из-за большой неточности их предсказания, однако стали фундаментом для создания модели Pareto/NBD, которая стала стандартом для моделирования поведения клиентов.

Ее первая версия была разработана Шмитлайном и Моррисоном в 1987 году для прогнозирования оттока клиентов [1]. С тех пор она была несколько модифицирована, в частности соединена с RFM (Recency, Frequency, Monetary) моделью [2, 3]. Ее суть заключается в описании потребительского поведения, когда клиент совершает заказы с постоянной интенсивностью в течение определенного периода времени, а затем становится неактивным. Если быть точнее, время выбытия моделируется с помощью Парето распределения (смесь гамма и экспоненциального распределений), в то время как потребительское поведение — с помощью отрицательного биномиального распределения (Пуассон-гамма).

В 2005 году Харди и Фейдер также разработали модель BG/NBD, представленную в качестве альтернативы существующей Pareto/NBD [4]. Обе модели показывают похожие результаты, однако в BG/NBD — намного более простой процесс оценки параметров. Между ними также есть и концептуальное различие — время выбытия в случае BG/NBD дискретно, т. е. клиент может «умереть» только после совершения заказа.

В 2011 году на основе моделировании оттока клиентов с помощью Pareto/NBD те же исследователи Харди и Фейдер предложили модель PDO (Periodic Death Opportunities), время выбытия в которой также дискретно, однако привязано к календарным дням, а не к конкретным сделкам [5]. Авторы модели утверждают, что она является предпочтительной для исследования оттока клиентов, тогда как для прогнозирования продаж лучшей остается Pareto/NBD.

2. Построение имитационной модели на основе Pareto/NBD

Модель Pareto/NBD была создана для предсказания покупательского поведения, которое имеет систематическую основу: каждый клиент покупает с определенной периодичностью, а затем становится неактивным. Стоит заметить, что модель предполагает случайный непрерывный процесс оттока клиентов. Если говорить более строго, то модель основывается на следующих пяти предпосылках:

- в период активности клиента количество его заказов в течение периода t распределено по закону Пуассона с интенсивностью транзакций λ;
- неоднородность в значениях интенсивности транзакций λ между клиентами подчиняется гамма-распределению с параметрами r и α;
- каждый клиент имеет ненаблюдаемый период жизни[2] длиной τ. Продолжительность жизни клиентов распределена экспоненциально с коэффициентом μ;
- неоднородность в значениях коэффициентов μ между клиентами подчиняется гамма-распределению с параметрами s и β;
- интенсивность транзакций λ и коэффициент μ независимы между собой и варьируют между клиентам.

Спецификация модели. Модель Pareto/NBD требует лишь три входные переменные, характеризующие прошлую покупательскую историю каждого клиента: количество повторных транзакций[3], сделанных в промежуток времени, на котором мы обучаемся, время между первым и последним заказом, и, собственно, длину этого промежутка. Система обозначений здесь следующая:

$$X = \left(x, t_x, T\right),$$

где x — количество транзакций, наблюдаемых в период (0, T]; $t_x\,(0 < t_x \le T)$ — время между первым и последним заказом.

В рамках данной модели получаются следующие ключевые результаты:

1. $\mathbf{E}[X(t)]$ — ожидаемое количество заказов, которое принесет случайный клиент в интервал времени (0, t].

2. $\mathbf{P}(alive \mid x, t_x, T)$ — вероятность, что клиент с наблюдаемым поведением (x, t_x, T) активен в момент времени T.

3. $\mathbf{E}[X(T, T+t) \mid x, t_x, T]$ — ожидаемое количество транзакций в конкретном периоде (T, $T + t$) для клиента с наблюдаемым поведением (x, t_x, T).

Также можно моделировать денежный поток, который принесут нам клиенты в конкретном интервале времени.

Генерация данных для имитационной модели. Реализация проведена в MATLAB, хотя эта модель может быть имплементирована в любой другой среде программирования, к примеру, под R существует целая библиотека для данной модели (BTYD package).

Порядок генерации данных следующий:

- априорный выбор значения для коэффициентов (r, α, s, β) для двух гамма-распределений переменных λ и μ;

[2] Под жизнью подразумевается период, когда клиент пользуется услугами компании, т. е. когда он активен.

[3] Далее везде будут подразумеваться именно повторные заказы, однако слово «повторные» в целях экономии пространства будет опущено.

– моделирование для каждого клиента значений параметров λ и μ на основе гамма-распределений с выбранными параметрами;
– имитация покупательской истории клиента (распределение заказов и время выбытия) исходя из полученных ранее значений параметров λ и μ.

Будем считать, что клиентская база состоит из 100 000 человек, которые делали заказы в промежутке времени длиной 104 недели (два года). Первый год обозначим как период обучения модели, второй же год будем считать тестовым. Покупательская история клиентов показана на рис. 1. Необходимо преобразовать ее в матрицу 100 000×3 в виде $X = (x, t_x, T)$. Пример матрицы приведен в табл. 1. Текст программного кода процесса моделирования приведен на рис. 2.

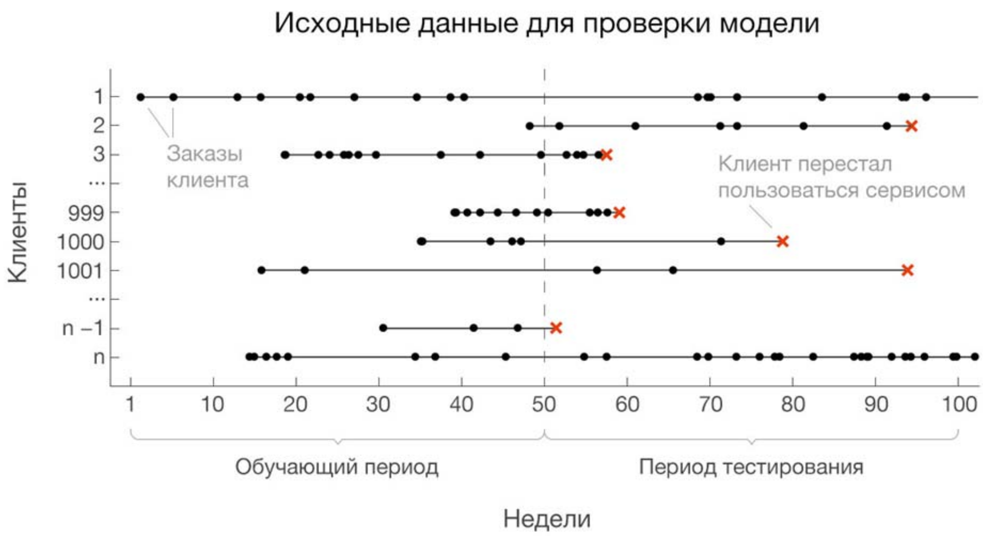

Рисунок 1. Покупательская история

Таблица 1. Пример входных данных для модели

	x	t_x	T
1	2	30.43	38.6
2	1	2.21	38.6
3	7	29.43	42.3
4	1	5.2	33.5

```
% Моделирование
clear; clc; format short;
[time_start, time_end] = deal(-52, 52);     % обучающий и тестовый периоды
n = 100000;                                      % количество клиентов
[r, alpha, s, beta] = deal(.8, 3, .2, .4);  % исходные параметры

lambda = gamrnd(r, 1./alpha, n, 1);     % интенсивность транзакций
mu = gamrnd(s, 1./beta, n, 1);          % интенсивность оттока клиентов
tau = exprnd(1./mu, n, 1);              % время выбытия

global x tx T par
T = rand(n, 1)*time_start - time_start; % время первого заказа
[x, tx, t, x2] = deal(zeros(n, 1));     % объявление переменных
isalive = ones(n, 1);                   % индикатор активности
% Моделирование периода обучения
for i = 1:n % цикл по всем клиентам
    j = 1;                              % первый заказ
    t(i, j) = exprnd(1./lambda(i));     % время между заказами
    isalive(i, j) = t(i, j) < tau(i);   % индикатор активности
    while isalive(i, j) ...         % пока пользователь "активен"
            && t(i, j) < T(i)       % и мы находимся в периоде обучения
        x(i) = x(i) + 1;        % количество повторных заказов
        tx(i) = t(i, j);        % время между первым и последним заказами
        j = j + 1;              % общее количество заказов
        t(i, j) = t(i, j - 1) + exprnd(1./lambda(i));   % следующий заказ
        isalive(i, j) = t(i, j) < tau(i);    % "активен" в момент времени t
    end
end
```

Рисунок 2. Текст програмы для моделирования

Оценка параметров модели. Четыре коэффициента модели Pareto/NBD (r, α, s, β) могут быть получены с помощью метода максимального правдоподобия. Общий вид логарифмической функции правдоподобия в нашем случае будет выглядеть следующим образом:

$$LL(r,\alpha,s,\beta)=\sum_{i=1}^{N}\ln\left[L\left(r,\alpha,s,\beta\mid X_i=x_i,t_{x_i},T_i\right)\right],$$

где

$$L\left(r,\alpha,s,\beta\mid X=x,t_x,T\right)=\frac{\Gamma(r+x)\alpha^r\beta^s}{\Gamma(r)}\left\{\frac{1}{(\alpha+T)^{r+x}(\beta+T)^s}+\left(\frac{s}{(r+s+x}\right)A_0\right\};$$

для $\alpha \geq \beta$:

$$A_0 = \frac{{}_2F_1\left(r+s+x, s+1; r+s+x+1; (\alpha-\beta/a+t_x)\right)}{(\alpha+t_x)^{r+s+x}} - \\ - \frac{{}_2F_1\left(r+s+x, s+1; r+s+x+1; \left(\alpha-\beta/\alpha+T\right)\right)}{(\alpha+T)^{r+s+x}};$$

для $\alpha < \beta$:

$$A_0 = \frac{{}_2F_1\left(r+s+x, r+x; r+s+x+1; \left(\beta-\alpha/\beta+t_x\right)\right)}{(\beta+t_x)^{r+s+x}} - \\ - \frac{{}_2F_1\left(r+s+x, r+x; r+s+x+1; \left(\beta-\alpha/\beta+T\right)\right)}{(\beta+T)^{r+s+x}}.$$

Здесь ${}_2F_1$ — гипергеометрическая функция Гаусса.

Реализация метода максимального правдоподобия для модели Pareto/NBD в MATLAB показана на рис. 3.

```
paramset = optimset('fmincon');
lb = 1e-6*ones(1, 4); ub = 100*ones(1, 4);  % нижняя и верхняя границы
par = ones(1, 4);                           % начальная точка
% Оптимизация логарифмической функции максимального правдоподобия
par = fmincon('PNBD_LL', par, [], [], [], [], lb, ub, [], paramset, [x tx T])
```

Рисунок 3. Использование метода максимального правдоподобия

Получив в итоге значения параметров модели (r, α, s, β), мы можем видеть, что они сходятся с теми, что мы задавали изначально (табл. 2). Как можно видеть, модель работает с приемлемой точностью.

Таблица 2. Сравнение начальных и оцененных значений параметров модели

	r	α	s	β
Начальные значения	0.8	3	0.2	0.4
Оцененные значения	0.8012	2.9981	0.2069	0.4392

Прогнозирование. В результате моделирования получаем возможность оценки трех ключевых параметров.

1. Ожидаемое количество заказов, которое принесет случайный клиент в интервал времени (0, t]:

$$\mathbf{E}\left[X(t)\,|\,r,\alpha,s,\beta\right]=\frac{r\beta}{\alpha(s-1)}\left[1-\left(\frac{\beta}{\beta+t}\right)^{s-1}\right].$$

2. Вероятность, что клиент с наблюдаемым поведением $X=(x,t_x,T)$ активен в момент времени T:

$$\mathbf{P}\left[active\,|\,r,\alpha,s,\beta,X\right]=\left\{1+\left(\frac{s}{r+s+x}\right)(\alpha+T)^{r+x}(\beta+T)^{s}A_0\right\}^{-1}.$$

3. Ожидаемое количество транзакций в конкретном периоде (T, $T+t$) для клиента с наблюдаемым поведением $X=(x,t_x,T)$:

$$\mathbf{E}\left[Y(t)\,|\,X,r,\alpha,s,\beta\right])=\frac{(r+x)(\beta+T)}{(\alpha+T)(s-1)}\left[1-\left(\frac{\beta+T}{\beta+T+t}\right)^{s-1}\right]\times\mathbf{P}\left[active\,|\,r,\alpha,s,\beta,X\right].$$

Приведенные на рис. 4, 5, графики иллюстрируют, насколько точно Pareto/NBD оценивает кумулятивное число заказов, а также их недельную динамику. На первом графике представлено, как Pareto/NBD моделирует динамику количества заказов в неделю выбранной когорты клиентов, негативный тренд которой объясняется тем, что со временем некоторые пользователи становятся неактивными. Можно сделать вывод об адекватности использования данной модели, так как, несмотря на вариацию реального количества заказов в неделю, прогнозные и действительные значения кумулятивного числа заказов практически совпадают (см. рис. 5). Данный вывод подтверждает и коэффициент детерминации регрессионной модели количества заказов в неделю, значение которого составило 0.86. Имплементация приведенных результатов в MATLAB представлена на рис. 6.

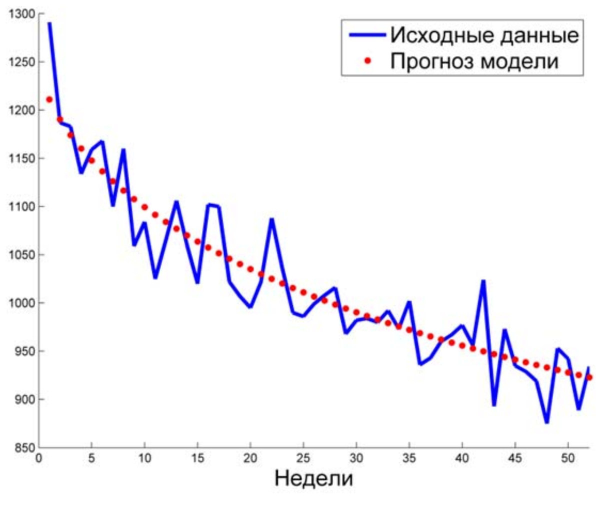

Рисунок 4. Количество заказов в неделю (сгенерированные данные)

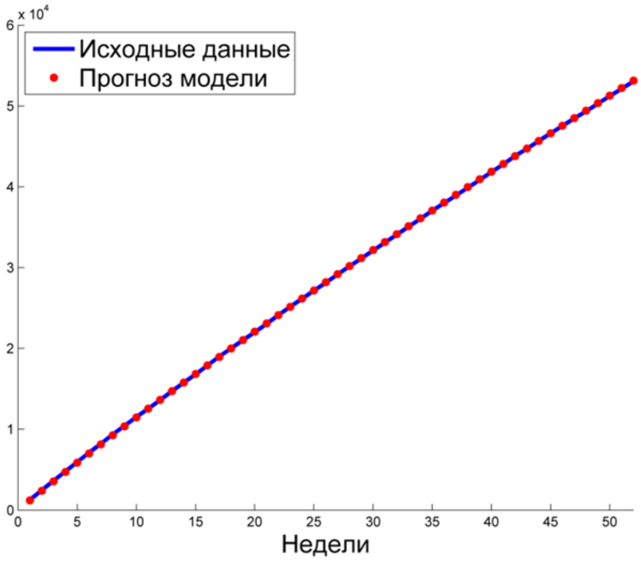

Рисунок 5. Кумулятивное количество заказов (сгенерированные данные)

```
% Моделирование тестового периода
for i = 1:n % цикл по всем клиентам
    j = x(i) + 1;                          % общее количество заказов
    while isalive(i, j) ...                % пока пользователь активен
            && t(i, j) < T(i) + time_end    % и в мы в тестовом периоде
        x2(i) = x2(i) + 1;      % количество заказов в тестовом периоде
        j = j + 1;
        t(i, j) = t(i, j - 1) + exprnd(1./lambda(i));   % следующий заказ
        isalive(i, j) = t(i, j) < tau(i);   % "активен" в момент времени t
    end
end
% Сравнение исходных данных с прогнозом модели
[real, est] = deal(zeros(time_end, 1));     % объявление переменных
A = repmat(T, 1, size(t, 2));
B = repmat(tau , 1, size(t, 2));
for i = 1:time_end  % цикл по неделям тестового периода
    real(i) = sum(sum(t > A + i - 1 & t < min(B, A + i), 2));
    est(i) = sum(PNBD_Est(x, tx, T, par, i - 1, i));    % прогноз модели

end
```

Рисунок 6. Реализация имитационной модели

```
% Построение графиков
close all; figure; hold on; % Количество заказов в неделю
plot(real, 'LineWidth', 3);                % исходные данные
plot(est, 'r.', 'MarkerSize', 16);         % прогноз модели
xlim([0 52]); xlabel('Недели', 'FontSize', 18)
legend({'Исходные данные', 'Прогноз модели'}, 'FontSize', 18);
print -painters -dtiff -r600 Orders.tiff

figure; hold on; % Общее количество заказов
plot(cumsum(real), 'LineWidth', 3);        % исходные данные
plot(cumsum(est), 'r.', 'MarkerSize', 18);  % прогноз модели
xlim([0 52]); xlabel('Недели', 'FontSize', 18)
legend({'Исходные данные', 'Прогноз модели'}, 'Location', 'NW', 'FontSize', 18);
print -painters -dtiff -r600 CumOrders.tiff
```

Рисунок 6 (продолжение). Реализация имитационной модели

3. Моделирование поведения клиентов компании Wheely

В данной главе будет произведена апробация модели Pareto/NBD. Для этого была взята обезличенная покупательская история клиентов компании Wheely одного из городов (анонимизированный ID клиента и дата совершения заказа), которые начали пользоваться услугами компании до февраля 2014 года, и приведена к виду подобно RFM, описанному выше в разделе спецификации. Выборка была разделена на обучающий и контрольный периоды. Обучающий период представляет собой время с начала основания компании до февраля 2014 года, когда наращивается база клиентов. Контрольный же период — это период тестирования нашей модели, т. е. прогнозирования поведения выбранной когорты потребителей.

В первую очередь, необходимо оценить параметры модели с помощью метода максимального правдоподобия. Весь процесс оптимизации был подробно описан в предыдущей главе. После того, как были оценены параметры, можно приступить к прогнозированию поведения клиентов в тестовом периоде. Имплементация модели для реальных данных приведена на рис. 7.

Чтобы понять, насколько хорошо модель справилась с поставленным заданием, проанализируем графики, приведенные на рис. 9, 10 (построение графиков — на рис. 8). Как можно видеть, модель адекватно аппроксимирует данные: прогноз общего количества заказов практически совпадает с реальными данными, а динамика количества заказов удовлетворительно описывается построенным трендом. Стоит все же отметить, что коэффициент детерминации регрессионной модели количества заказов в неделю составляет всего 0.3, однако столь низкое значение во многом можно объяснить значительным спадом, который пришелся на майские праздники. Можно сделать вывод о том, что, несмотря на существующие неточности, модель Pareto/NBD является приемлемой для нашего анализа, так как, по крайней мере, не наблюдается очевидных противоречий ее реальным данным.

```
clear; load Wheely
t0 = datenum(2014, 2, 1);        % начало прогнозируемого периода
Train = D(D.ordered_at < t0, :);                            % обучающий период
Test = D(D.ordered_at >= t0 & D.user_created_at < t0, :);   % тестовый период
U = grpstats(Train, 'id', {'min', 'max'}, 'DataVars', 'ordered_at');

global x tx T par
x = U.GroupCount - 1;            % количество повторных заказов
% время между первым и последним заказами в неделях
tx = (U.max_ordered_at - U.min_ordered_at)/7;
T = (t0 - U.min_ordered_at)/7;   % время с момента первого заказа

paramset = optimset('fmincon');
lb = 1e-6*ones(1, 4); ub = 100*ones(1, 4);  % нижняя и верхняя границы
par = ones(1, 4);                           % начальная точка
% Оптимизация логарифмической функции максимального правдоподобия
par = fmincon('PNBD_LL', par, [], [], [], [], lb, ub, [], paramset, [x tx T])

for i = 1:30 % цикл по неделям тестового периода
    t(i) = t0 + 7*(i - 1);  % недели в числовом формате
    real(i) = sum(Test.ordered_at >= t(i) & Test.ordered_at < t(i) + 7);
    est(i) = sum(PNBD_Est(x, tx, T, par, i - 1, i));
end
```

Рисунок 7. Применение модели для реальных данных

```
close all; figure; hold on; % Количество заказов в неделю
plot(t, real, 'LineWidth', 3); xlabel('Месяц', 'FontSize', 18)
plot(t, est, 'r.', 'MarkerSize', 16); datetick('x', 'mmm')
legend({'Наши данные', 'Прогноз модели'}, 'FontSize', 18);
print -painters -dtiff -r600 WheelyOrders.tiff

figure; hold on; % Общее количество заказов
plot(t, cumsum(real), 'LineWidth', 3); xlabel('Месяц', 'FontSize', 18)
plot(t, cumsum(est), 'r.', 'MarkerSize', 16); datetick('x', 'mmm')
legend({'Наши данные', 'Прогноз модели'}, 'Location', 'NW', 'FontSize', 18);

print -painters -dtiff -r600 WheelyCumOrders.tiff
```

Рисунок 8. Построение графиков

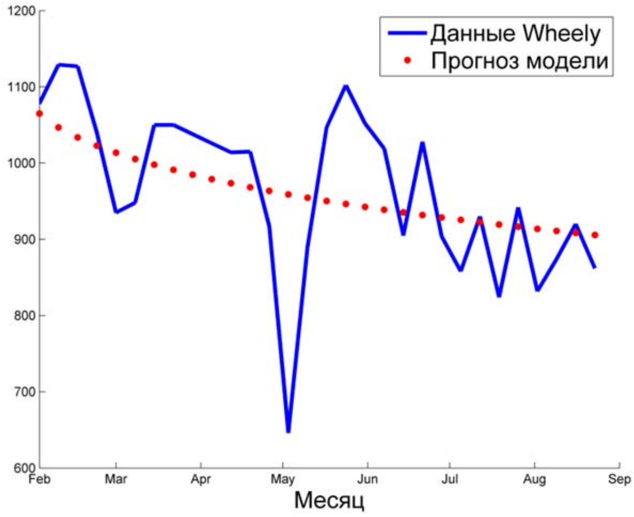

Рисунок 9. Количество заказов в неделю (реальные данные)

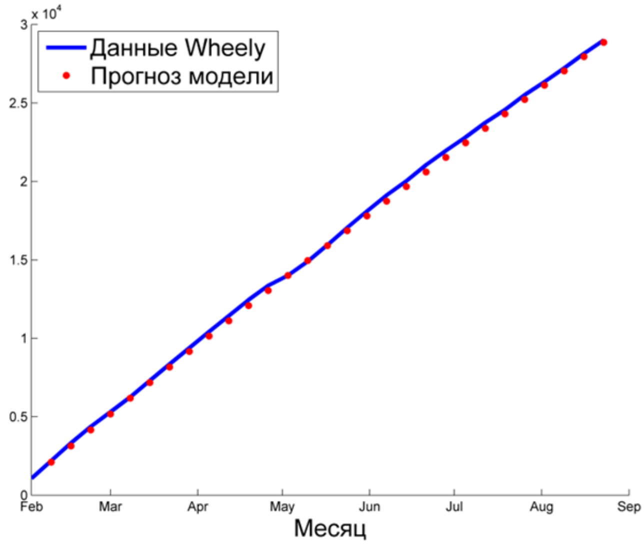

Рисунок 10. Кумулятивное количество заказов (реальные данные)

4. Результаты и обсуждение

Модель Pareto/NBD позволяет узнать: вероятность, что пользователь стал неактивным; количество заказов, которые принесет случайный пользователь; количество заказов, которое принесет определенная когорта клиентов. Успешная реализация модели на реальных данных Wheely доказывает, что модель действительно пригодна для решения прикладных бизнес-задач. Обладание такой информацией дает возможность:

- предотвращать отток клиентов, высылая специальные предложения тем, кто, согласно модели, находится на грани «выбытия», или же пытаться вернуть неактивных;
- прогнозировать оборот, который мы получим от определенной когорты потребителей;
- оптимизировать затраты на привлечение новых клиентов, исходя из информации о том, сколько заказов принесет случайный пользователь.

Таким образом, проведенные исследования служат доказательством того, что модель Pareto/NBD является мощным инструментом для клиентской аналитики и до сих пор не потеряла своей актуальности.

Литература

[1] *Schmittlein D. C., Morrison D. G., Colombo R.* Counting your customers: Who are they and what will they do next? // Management Science. 1987. Vol. 33. No. 1. P. 1–24.

[2] *Fader P. S., Hardie B. G. S.* A Note on Deriving the Pareto/NBD Model and Related Expressions 2005 (http://www.brucehardie.com/notes/009/pareto_nbd_derivations_2005-11-05.pdf)

[3] *Fader P. S., Hardie B. G. S., Lee K. L.* “Counting your customers” the easy way: An alternative to the Pareto/NBD model // Marketing Science. 2005. Vol. 24. No. 2. P. 275–284.

[4] *Jerath K., Fader P., Hardie B.* New Perspectives on Customer ‘Death’ Using a Generalization of the Pareto/NBD Model // Marketing Science. 2011. Vol. 30. No. 5. P. 866–880.

[5] *Fader P. S., Hardie B.* Using Iso-Value Curves for Customer Base Analysis // Journal of Marketing Research. 2005. Vol. 42. P. 135–144.

Авторы:

Озеров Александр Георгиевич, бизнес-аналитик компании Wheely

Соловьев Роман Андреевич, бизнес-аналитик компании Wheely

Приложения

Логарифмическая функция максимального правдоподобия

```
function f = PNBD_LL(params, data)
r     = params(1);
alpha = params(2);
s     = params(3);
beta  = params(4);
x = data(:,1);
tx = data(:,2);
T = data(:,3);

maxab = max(alpha, beta);
absab = abs(alpha - beta);
param2 = s + 1;
if alpha < beta
    param2 = r + x;
end
z1 = absab./(maxab + tx);
z2 = absab./(maxab + T);
a = r + s + x;
b = param2;
c = r + s + x + 1;
A = r*log(alpha) + s*log(beta) - gammaln(r) + gammaln(r + x);
B = 1./((alpha + T).^(r + x).*(beta + T).^s);
if absab == 0
    F1 = 1./((maxab + tx).^(r + s + x));
    F2 = 1./((maxab + T).^(r + s + x));
else
    F1 = H2F1(a, b, c, z1)./((maxab + tx).^(r + s + x));
    F2 = H2F1(a, b, c, z2)./((maxab + T).^(r + s + x));
end
C = (s./(r + s + x)).*(F1 - F2);
Log_likelihood = A + log(B + C + realmin);

f = -sum(Log_likelihood);
end
```

Гипергеометрическая функция. Существует множество алгоритмов для вычисления гипергеометрической функции, которые сильно отличаются по скорости работы и точности вычислений. В данном примере представлена наиболее простая и достаточно быстрая в реализации процедура. В этой функции не проверяются некоторые граничные условия, но для нашей модели это не столь существенно.

```
function f = H2F1(a, b, c, z)

lenz = length(z);
j = 0;
uj = ones(lenz, 1);
f = uj;
lteps = 0;

while lteps < lenz
   lasty = f;
   j = j + 1;
   uj = uj.*(a + j - 1).*(b + j - 1)./(c + j - 1).*z./j;
   f = f + uj;
   lteps = sum(f == lasty);
end

end
```

Ожидаемое количество транзакций в конкретном периоде (T, $T + t$) для клиента с наблюдаемым поведением (x, t_x, T).

```
function f = PNBD_Est(x, tx, T, params, time_start, time_end)
r     = params(1);
alpha = params(2);
s     = params(3);
beta  = params(4);

time_end = time_end.*(time_start + T > 0);
time_start = time_start.*(time_start + T > 0);

tmp = (r + x).*(beta + T)./((alpha + T).*(s - 1));
tmp_start = ((beta + T)./(beta + T + time_start)).^(s - 1);

tmp_end = ((beta + T)./(beta + T + time_end)).^(s - 1);
nbr_start = tmp.*(1 - tmp_start);
nbr_end = tmp.*(1 - tmp_end);

Palive = PNBD_Palive(x, tx, T, params);

f = (nbr_end - nbr_start).*Palive;
end
```

Вероятность, что клиент с наблюдаемым поведением (x, t_x, T) активен в момент времени T.

```
function f = PNBD_Palive(x, tx, T, params)
r     = params(1);
alpha = params(2);
s     = params(3);
beta  = params(4);

maxab = max(alpha, beta);
absab = abs(alpha - beta);
par = s + 1;
if alpha < beta
    par = r + x;
end
A1 = ((alpha + T)./(maxab + tx)).^(r + x).*((beta + T)./(maxab + tx)).^s;
B1 = ((alpha + T)./(maxab + T)).^(r + x).*((beta + T)./(maxab + T)).^s;
A2 = H2F1(r + s + x, par, r + s + x + 1, absab./(maxab + tx));
B2 = H2F1(r + s + x, par, r + s + x + 1, absab./(maxab + T));

f = 1./(1 + (s./(r + s + x)).*(A1.*A2 - B1.*B2));
end
```

Ожидаемое количество заказов, которое принесет случайный клиент в интервал времени (0, t].

```
function f = PNBD_E(params, t)
r     = params(1);
alpha = params(2);
s     = params(3);
beta  = params(4);

A = r*beta/alpha/(s - 1);
B = 1 - (beta/(beta + t))^(s - 1);

f = A*B;
end
```

Customer behaviour prediction: the case of Wheely

Alexander Ozerov, Roman Solov'ev

*Wheely (*http://wheely.com*)*
117105, Warsawskoe highway, 1, 1, Moscow, Russia
e-mail: ao@wheely.com, roman@wheely.com

Abstract. This paper is designed to describe the process of customer behaviour prediction in companies with continuous and noncontractual transactions. The work is based on the case of personal driver service Wheely which anonymised client data is used for the research analysis. The paper shows how the company is solving a problem of customer churn regulation using Pareto / NBD model. It is also explained how the results of this model implementation allows to improve the quality and efficiency of the service.
Keywords: Pareto / NBD, customer churn, customer lifetime value, Wheely.

Reference

[1] *Schmittlein D. C., Morrison D. G., Colombo R.* (1987) Counting your customers: Who are they and what will they do next? *Management Science*, 33(1), 1–24.

[2] *Fader P. S., Hardie B. G. S.* (2005) A Note on Deriving the Pareto/NBD Model and Related Expressions (http://www.brucehardie.com/notes/009/pareto_nbd_derivations_2005-11-05.pdf)

[3] *Fader P. S., Hardie B. G. S., Lee K. L.* (2005) "Counting your customers" the easy way: An alternative to the Pareto/NBD model. *Marketing Science,* 24(2), 275–284.

[4] *Jerath K., Fader P., Hardie B.* (2011) New Perspectives on Customer 'Death' Using a Generalization of the Pareto/NBD Model. *Marketing Science*, 30(5), 866–880.

[5] *Fader P. S., Hardie B.* (2005) Using Iso-Value Curves for Customer Base Analysis. *Journal of Marketing Research*, 42, 135–144.