

Дюличева Ю.Ю., Рябченко Е.А.

УДК 519.865

О ПОДХОДАХ К МОДЕЛИРОВАНИЮ И ОПТИМИЗАЦИИ РАБОТЫ С ПОТРЕБИТЕЛЯМИ УСЛУГ

Аннотация. В статье анализируются этапы моделирования поведения клиентов (потребителей услуг), включающие фильтрацию (очистку) клиентской базы, сегментирование, оценку рисков, связанных с отказом клиента от услуг и товаров, оптимизацию работы с клиентами; исследуются особенности применения методов Data Mining, в частности, решающих деревьев и решающих списков, методов кластерного анализа, генетических алгоритмов, байесовских сетей и предлагаются перспективные подходы к моделированию поведения клиентов; приводятся примеры CRM-систем.

Ключевые слова: CRM-система, RFM-анализ, Data Mining, сегментирование клиентов.

Анотація. У статті аналізуються етапи моделювання поведінки клієнтів такі, як фільтрація (очищення) клієнтської бази, сегментування клієнтів, оцінювання ризиків, що пов'язані з відмовою клієнтів від товарів та послуг, оптимізація роботи з клієнтами; досліджуються особливості застосування методів Data Mining, а саме розв'язуючих дерев і розв'язуючих списків, нейронних мереж, генетичних алгоритмів, байєсовських мереж і пропонуються перспективні напрями до моделювання поведінки клієнтів; наводяться приклади CRM-систем.

Ключові слова: CRM-система, RFM-аналіз, Data Mining, сегментування клієнтів.

Summary. The analysis of the modeling of the customer behavior stages such as filtering (cleaning methods) of the client data base, customer segmentation, risk evaluation based on customer churn prediction, customer relationship optimization is considered in the paper.

The first stage of the customer behavior understanding is the client data base preprocessing including data cleaning and data filtering. The important part of the complex data cleaning methods is outlier detection. The definitions of a clear outlier and a majority outlier are proposed. The examples of the filtering and cleaning data methods realized in SQL Server 2012 platform, IBM SPSS, analytical platform Deductor are discussed.

The second stage of the customer behavior understanding is customer segmentation. The perspectives of the combined customer segmentation methods based on the RFM-analysis and Data Mining are investigated. The examples of the client segmentation based on IBM SPSS and analytical platform Deductor are considered.

The third stage of the customer behavior understanding is customer churn prediction models. The possibilities of the survival analysis for customer behavior are considered. The examples of the survival analysis based on Cox Regression and Kaplan-Meier procedure realized in SPSS are considered.

The fourth stage of the customer behavior understanding is fraud and anomaly detection. The perspectives of outlier detection methods for the fraud recognition are discussed.

The fifth stage of the customer behavior understanding is customer preferences prediction methods and usage of CRM-systems. The perspectives of Data Mining techniques for the customer preferences prediction and examples of the well-known CRM-systems are considered.

Key words: CRM-system, RFM-analysis, Data Mining, customer segmentation.

Постановка проблемы. Прогнозирование прибыли и планирование развития бизнеса невозможно без участия клиентов и эффективного управления взаимоотношениями с клиентами (Customer Relationship Management (CRM)). Клиентские базы данных могут содержать информацию о держателях дисконтных карт, посетителях супермаркетов, пользователей интернет-услуг, посетителях интернет-магазинов, клиентах банка и т.д. Правило Парето гласит, что «20% клиентов приносят 80% прибыли», известно также, что гораздо выгоднее удержать клиента, чем искать нового, поэтому для компаний анализ и прогнозирование поведения клиентов, разработка обоснованной программы лояльности как способа удержания клиентов; автоматизация процесса управления взаимоотношениями с клиентами (разработка эффективных CRM-систем) являются важнейшими задачами бизнес-аналитики. Изучение поведения клиентов и понимание их потребностей позволяют не только обоснованно планировать маркетинговые акции на основе выявления предпочтений клиентов, но и прогнозировать прибыль, основываясь на целевой аудитории, а также на основе шаблонов типичного поведения клиентов выявлять мошенников (fraud detection), поэтому поиск эффективных моделей прогнозирования и оптимизации работы с клиентами (потребителями товаров и услуг) является актуальной задачей.

Целью статьи является анализ подходов к моделированию и оптимизации работы с клиентами на основе методов Data Mining.

Анализ последних достижений и публикаций. Специфичность задачи анализа клиентских баз данных и заинтересованность компаний в разработке и внедрении эффективных систем управления взаимоотношениями с клиентами привлекает большое число зарубежных и отечественных ученых к исследованию этой задачи. Так, вопросы сбора информации и формирования клиентской базы рассматриваются в работе Кендра Ли [1]. Исследованием сегментирования клиентских баз, оценкой оптимального числа и качества сегментов занимались Чен (Chan), Чунг (Chung), Джонс (Jones), Ву (Woo), Хванг (Hwang), Джунг (Jung), Зенг (Zhang) и т.д. [2] Понятие клиентской среды и один из алгоритмов анализа клиентских сред предложен в работе Воронцова К.В. и Лексина В.А. [3]

Изложение основного материала. Выделим и проанализируем основные этапы решения задачи моделирования поведения клиентов.

1. Создание клиентской базы и фильтрация (очистка) данных о клиентах. В процессе взаимодействия с клиентами происходит накопление огромных объемов данных, анализируя которые можно больше узнать о своих клиентах, выявить закономерности в их поведении. Однако данные часто содержат ошибки,

неточности (выбросы), которые затрудняют выявление устойчивых закономерностей. Очистка данных позволяет устранить выбросы. Определения чистого и мажоритарного выбросов были предложены автором в работе [4]. Пусть X^* – обучающая выборка, $y^*: X \rightarrow Y$ – неизвестная целевая зависимость, X – множество всех допустимых объектов, Y – множество ответов (меток классов). Известны значения целевой зависимости y^* на обучающей выборке, т.е. $y_i = y^*(x_i)$, $i=1 \dots \lambda$. Пусть $L(a(\tilde{X}), y^*(\tilde{X})) = [a(\tilde{X}) \neq y^*(\tilde{X})]$ – функция потерь от ошибки при классификации обучающего объекта \tilde{X} алгоритмом $a \in A$, где A – некоторое множество алгоритмов обучения. Объект $\tilde{X} \in X^*$ называется чистым выбросом относительно множества A алгоритмов обучения, если $L(a(\tilde{X}), y^*(\tilde{X})) = 1$, для любого алгоритма $a \in A$. Обучающий объект \tilde{X} называется мажоритарным выбросом относительно множества A алгоритмов обучения, если существуют такие $m > \lfloor |A|/2 \rfloor$ алгоритмов обучения $a \in A$, для каждого из которых выполняется условие $L(a(\tilde{X}), y^*(\tilde{X})) = 1$.

Рассмотрим некоторые программные продукты (системы, платформы), имеющие встроенные средства для фильтрации и очистки данных. Встроенный фильтр данных, а также средства интеллектуального анализа данных поддерживаются в платформе SQL Server 2012.

IBM SPSS (Statistical Package for the Social Sciences) Modeler – рабочая среда с визуальным интерфейсом и средствами аналитики для сбора и анализа данных, содержащая модули предобработки данных для выявления выбросов, аномальных данных (Anomaly Detection) [5].

Система Deductog содержит модуль обработки данных, позволяющий производить фильтрацию и очистку данных на основании статистических функций и алгоритмов Data Mining [6].

2. Сегментирование клиентов (выделение групп клиентов со схожими предпочтениями на основе самоорганизующихся карт или алгоритмов кластерного анализа). Использование различных подходов к сегментированию клиентской базы позволяет решить следующие важные задачи: выделить группы клиентов с высокой частотой приобретения товаров или услуг; выявить группы постоянных клиентов; выявить группы наиболее значимых (выгодных) клиентов, приобретающих товары и услуги на значительные суммы; выявить сегменты перспективных будущих клиентов (например, полностью неопределенных с выбором товаров и услуг), а также клиентов, приобретающих товары и услуги с определенной периодичностью. Такой всесторонний анализ сегментов клиентской базы позволяет разрабатывать особые программы лояльности, специализированные сервисы и другие маркетинговые мероприятия, которые будут результативными именно для данного сегмента клиентской базы. Сегментирование позволяет также получать профили клиентов (customer profile) – набор характеристик, присущих определенному сегменту клиентов. Часто для сегментирования клиентов используется RFM-анализ (или его модификация RFD, учитывающая общую продолжительность знакомства с клиентами). RFM-анализ (Recency, Frequency, Monetary) основан на использовании трех величин: R (Recency) – разность в днях между текущей датой и датой последнего заказа; F (Frequency) – общая частота заказов клиентов, т.е. общее количество заказов за все время взаимодействия с клиентом; M (Monetary) – общая сумма средств, потраченных клиентом за время сотрудничества с ним. На основании значений каждой из трех величин клиенты разбиваются на 5 групп, каждая из которых получает значение от 1 до 5: значение 1 получает наименее вероятная группа, значение 5 – наиболее вероятная группа. К недостаткам метода можно отнести нефиксированные границы отсечения по вероятности отклика, поэтому были предложены различные модификации использования RFM, например, модификация RFM с учетом поведения клиентов на различных этапах жизненного цикла [7].

Хорошо зарекомендовали себя для сегментирования клиентских баз алгоритмы Data Mining. Для сегментирования клиентских баз часто используются такие методы кластеризации как метод k-средних, двухшаговый метод кластерного анализа (TwoStep), самоорганизующиеся карты Кохонена. Модули реализации сегментирования данных, с помощью перечисленные выше методов, реализованы в среде IBM SPSS Modeler [5]. Для извлечения закономерностей из клиентских баз широко используются решающие деревья и решающие списки. Решающие деревья в таких системах, как правило, реализованы на основании алгоритмов C5.0 (C 4.5) и CHAID. Пример сегментирования клиентов на основе дерева решений и его реализация на основе аналитической платформы Deductog представлен на рисунке 1.

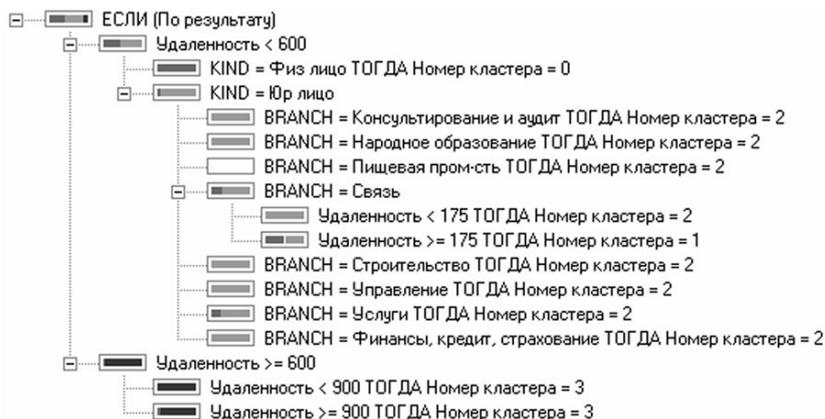


Рис. 1. Выделение сегментов в системе Deductog [6] на основе дерева решений

Для поиска и выявления целевых сегментов клиентов используются сложные комбинированные подходы, основанные на RFM и LTV (Life Time Value – оценка жизненного цикла клиента) анализе в сочетании с методами Data Mining. Такой подход позволяет сочетать методики выявления сегментов из клиентской базы, учитывая как поведенческие особенности групп клиентов, так и скрытые закономерности в их поведении.

3. Оценивание вероятности отказа клиентов от услуг (оценивание риска оттока клиентов). При оценке рисков, связанных с отказом клиентов от услуг или товаров, используется анализ выживаемости (survival analysis). Анализ “выживаемости” позволяет оценить вероятность того, что клиент компании, сотрудничавшей с ней некоторый период времени, продолжит пользоваться услугами или товарами компании в следующий период времени или оценивание вероятности того, что клиент, сотрудничавший с компанией определенный период времени, решит отказаться от ее товаров и услуг.

Анализ “выживаемости” клиентов можно выполнить в программе STATISTICA, оценивая вероятности, характеризующие “выживаемость” клиентов, частотами, выделяя группы клиентов, с которыми связаны наибольшие риски отказа от товаров и услуг, и используя модель Кокса с зависящими от времени ковариатами.

Еще один программный продукт SPSS предоставляет возможность проведения анализа “выживаемости” клиентов на основании таблиц выживаемости, процедуры Каплана-Мейера (Kaplan-Meier procedure) и регрессии Кокса или модели пропорциональных рисков, возвращающей в качестве результата отношение рисков (hazard ratio) и его доверительный интервал. Для выполнения анализа “выживаемости” на основе перечисленных моделей следует выбрать пункт меню Analyze->Survival->Kaplan-Meier или Analyze->Survival->Cox Regression [5]. Пример анализа “выживаемости” на основе процедуры Каплана-Мейера в программе SPSS и построение соответствующих кривых для групп клиентов приведен на рисунке 2.

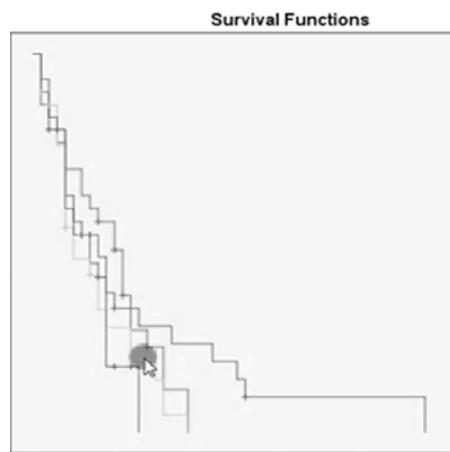


Рис. 2. Выполнение анализа “выживаемости” клиентов на основе процедуры Каплана-Мейера в программе SPSS

Перспективным направлением для получения критериев “выживаемости” клиентов является применение алгоритмов Data Mining, в частности, применение нечеткой логики и байесовских сетей.

4. Выявление мошенников является одним из направлений исследования поведения клиентов. Очевидным способом выявления мошенников является разработка шаблонов типичного поведения клиентов на основе закономерностей, характерных для сегментов клиентов. Еще один способ выявления мошенников, связанный с описанными выше этапами моделирования поведения клиентов – построение кривых, характеризующих выживаемость клиентов и оценивание рисков, связанных с отказом клиентов от товаров и услуг. В настоящее время выявление мошенничества (fraud detection) становится одним из важнейших направлений развития систем бизнес-аналитики, поскольку мошенники способны нанести существенные убытки компании. Выявление подозрительных действий или махинаций при выполнении операций над большими объемами данных является сложной задачей, для решения которой применяются методы Data Mining, направленные на обучение (обобщение) и распознавание подозрительных операций. В частности, для решения задачи обнаружения мошенничества используются байесовские сети, нейронные сети, деревья решений и деревья регрессии, бинарные диаграммы решений, SVM.

Одним из программных продуктов, реализующих задачу выявления мошенничества, является аналитическая система Deductor, имеющая модуль выявления подозрительных действий на основе следующих этапов: обнаружение отклонения от технологии, выявление аномалий и сегментирование по операциям и сотрудникам [6].

Перспективным направлением для выявления мошенничества является выявление выбросов во множестве операций и выявление выбросов на множестве клиентов. Выброс характеризует нетипичное поведение клиента или нетипичную операцию, для которой невозможно построить обобщающую закономерность. Решение задачи выявления мошенничества на основе анализа выбросов (Filter Outliers) реализовано в программных продуктах SAS [8].

5. Оптимизация работы с клиентами. Одним из направлений оптимизации работы с клиентами является разработка эффективных алгоритмов прогнозирования предпочтений клиентов. На основании мнений клиентов о предоставляемых товарах и услугах требуется предсказать ассортимент, который наиболее вероятно предпочтет клиент. Для описания мнений клиентов удобно использовать лингвистические переменные, а для составления прогноза такие методы Data Mining как нейронные сети и генетические алгоритмы [9]. Прогнозирование предпочтений клиентов может выполняться на основании выявленных сегментов клиентов и характеризующих их закономерностях.

Еще одним направлением оптимизации работы с клиентами является внедрение автоматизированных систем управления взаимоотношениями с клиентами (CRM-систем). В последнее время на рынке наблюдается увеличение числа программных продуктов для автоматизации проведения клиентской аналитики и анализа клиентских сред. Перечислим некоторые из них: 1С:CRM; Deductor:CRM; STATISTICA; веб-ориентированная CRM: Клиентская база; библиотека для всестороннего анализа клиентской среды Forecsys MetricsPro; CRM “Битрикс 24” и другие.

Выводы. В настоящее время предложено больше число подходов к анализу клиентских баз на основе методов Data Mining. Однако для эффективного использования таких методов при выполнении анализа реальных клиентских баз требуется правильно подбирать настройки параметров алгоритмов, например, при использовании метода k-средних требуется правильная оценка числа сегментов в клиентской базе; при построении решающего дерева по методу CHAID требуется указать число уровней/глубину дерева, минимальное количество объектов (описаний клиентов) в дочерних ветвях, параметры для ветвления и слияния и т.д..

Поиск и разработка эффективных алгоритмов и их реализация в системах управления взаимоотношениями с клиентами позволяет оптимизировать работу с клиентами по всем направлениям и, прежде всего, проводить обоснованные маркетинговые мероприятия, предназначенные, согласно выявленным предпочтениям, непосредственно для целевых сегментов клиентской базы, т.е. выполнять рассылку рекламной информации не хаотично, а целенаправленно для сегментов клиентов, которых заинтересует именно эта информация и будет с большой вероятностью иметь отклик.

Перспективным направлением для дальнейших исследований является поиск комбинированных алгоритмов, направленных на всесторонний анализ клиентских баз данных и разработка критериев оценки выгодных клиентов с учетом особенностей поведения клиентов.

Источники и литература:

1. Кендра Ли. Создание клиентской базы : пошаговое руководство по превращению контактов в деньги / Кендра Ли ; пер. с англ. Болдышева А. В. – М. : Вершина, 2006. – 360 с.
2. Chan C. C. H. Intelligent value-based customer segmentation method for campaign management : A case study of automobile retailer / C. C. H. Chan // Expert Systems with Applications. – 2008. – 34. – pp. 2754-2762.
3. Воронцов К. В. Анализ клиентских сред: выявление скрытых профилей и оценивание сходства клиентов и ресурсов / К. В. Воронцов, В. А. Лексин // Сборник тезисов 16-й Всероссийской конференции «Математические методы распознавания образов-2013». – 2013.
4. Дюличева Ю. Ю. О задачах фильтрации обучающих данных / Ю. Ю. Дюличева // Искусственный интеллект. – 2006. – №2. – С. 65-71.
5. IBM SPSS [Электронный ресурс]. – 25. 07. 2013. – Режим доступа : <http://www-01.ibm.com/software/analytics/spss>
6. Аналитическая платформа Deductor [Электронный ресурс]. – Режим доступа : www.basegroup.ru
7. Полежаев И. Е. Метод сегментации клиентских баз данных на основе жизненного цикла клиента // Исследовано в России. – 2006 [Электронный ресурс]. – 25.07.2013. – Режим доступа : <http://zhurnal.ape.relarn.ru/articles/2006/200.pdf>
8. Аналитика поведения клиентов на основе продуктов SAS. [Электронный ресурс]. – 25.07.2013. – Режим доступа : <http://www.sas.com/events/aconf/2010/bdmci61.html>
9. Mashinchi M. R. Constructing a Customer's Satisfactory Evaluator System Using GA-Based Fuzzy Artificial Neural Networks / M.Reza Mashinchi, Ali Selamat // World Applied Sciences Journal, 5(4). – 2008. – pp.432-440.