

МОДЕЛЬ ЗАДАЧИ КЛАССИФИКАЦИИ КЛИЕНТОВ ПО СТАДИЯМ ВЗАИМООТНОШЕНИЙ В СОВРЕМЕННЫХ КОРПОРАТИВНЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

Рассматривается модель задачи управления взаимоотношениями с клиентами как задачи оценки стабильности продаж. Предлагается методика построения системы поддержки принятия решений об основных параметрах задачи – модели тренда, периода для анализа, числа таких периодов. На реальных данных демонстрируется эффективность предлагаемой методики с точки зрения перспектив её реализации в современных корпоративных информационных системах.

Введение

Под корпоративными информационными системами (КИС) в настоящее время принято понимать совокупность средств сбора, передачи, первичной обработки и хранения данных, сопровождающих деятельность предприятий и организаций, структурирования и анализа этих данных и выработки набора возможных управленческих действий. Основной задачей современных корпоративных систем является создание информационной базы для принятия управленческих решений руководством предприятий, при этом содержание и форма представления информации должны обеспечивать принятие как можно меньшего числа ошибочных решений. КИС всегда существуют в рамках некоторой бизнес-модели (бизнес-процесса) организации².

¹ Гриняк Виктор Михайлович, канд. техн. наук, доцент кафедры ИСКТ ВГУЭС. Viktor.Grinyak@vvsu.ru

Петухов Михаил Ильич, аспирант кафедры ИСКТ ВГУЭС. Научный руководитель – Гриняк Виктор Михайлович

Специальность 05.13.18 – Математическое моделирование, численные методы и комплексы программ

² Жданов Б.Н. Новая логика и факторы развития КИС // Корпоративные системы. 2003. №3.

Функциональность современных КИС включает в себя решение таких задач как стратегическое и оперативное планирование, товарный и бухгалтерский учёт, управленческий учёт по различным направлениям. Технологически перечисленные задачи поддерживаются программными модулями с названиями типа: «Планирование и управление финансами», «Управление персоналом», «Модуль бухгалтерского и налогового учёта», «Логистика», «Планирование и управление производством», «Управление автотранспортом» и др.¹

Управление взаимоотношениями с клиентами (Client Resource Management, CRM) является одной из важнейших составляющих современного управленческого учёта. Суть задачи CRM состоит в накоплении информации о клиентах, которая может использоваться при проведении маркетинговых исследований, формировании политики предприятия по отношению к тому или иному клиенту и т.п. Методологически функциональность CRM предполагает ведение каталогов объектов базы данных – клиентов (дилеров, партнёров, поставщиков, покупателей, конкурентов), формирование некоторого множества пользовательских классификаторов и атрибутов, присваивание их клиентам и отображение каталогов клиентов с группировкой их по выбранному набору классификационных признаков. Подсистемы CRM того или иного содержания имеются практически во всех современных корпоративных информационных системах, получивших широкое распространение на отечественном рынке (такие, например, как российские «Флагман», «1С», «Галактика» и зарубежные «Ахарт», «SAP» и др.)².

Одной из задач, часто реализуемых в рамках функциональности CRM, является задача классификации клиентов (как правило – покупателей) по стадиям взаимоотношений. Стадии могут быть, например, следующие: потенциальный покупатель, разовый покупатель, постоянный покупатель, потерянный покупатель. С точки зрения управленческой деятельности, результаты решения задачи классификации клиентов по стадиям взаимоотношений могут быть использованы для анализа значимости клиентов, их надёжности, стабильности, анализа эффективности работы менеджеров, априорной достоверности планирования.

Классификация стадий взаимоотношений с покупателями основана на ретроспективном анализе результатов взаимоотношений за некоторый период; анализируемым показателем может выступать, например, количество сделок, объём продаж, выручка от продаж, при-

¹ Бочаров Е.П., Колдина А.И. Интегрированные корпоративные информационные системы. – М.: Финансы и статистика, 2005.

Turban E., McLean E., Wetherbe J. Information Technology for Management: Transforming Business in the Digital Economy. – New York: John Wiley & Sons, 2002.

² Бочаров Е.П., Колдина А.И. Интегрированные корпоративные информационные системы. – М.: Финансы и статистика, 2005.

Turban E., McLean E., Wetherbe J. Information Technology for Management: Transforming Business in the Digital Economy. – New York: John Wiley & Sons, 2002.

быть от продаж и т.п. При этом все современные КИС предоставляют возможность пользователю выбирать параметры решения задачи (анализируемый показатель, период анализа, число периодов для анализа) исключительно интуитивно, «вручную».

В настоящей работе рассматривается модель задачи классификации клиентов по стадиям взаимоотношений и предлагается метод автоматизированного выбора её параметров на основе результатов статистического анализа.

Основные модельные представления

Как правило, в задачах экономического анализа минимальным периодом оценивания какого-либо показателя являются сутки, соотнесённые с конкретной датой; часто используются и другие периоды: «недельный», «месячный», «квартальный» и т.д. Значение выбранного показателя X_k в период с номером k может быть выражено формулой

$$X_k = \sum_{i=(k-1)*n}^{k*n} x_i, \quad (1)$$

где n – число дней в периоде, за который проводится анализ, x_i – значение выбранного показателя в день с номером i , $i = \overline{1, N}$, $k = \overline{1, J}$, N – общее число дней, за которые взяты данные, J – число периодов, на которые разбиты N дней, так что $N = Jn$.

Модель изменения значения показателя X во времени может быть выражена формулой

$$X_k = G(k) + \eta(k), \quad (2)$$

где $G(k)$ – функция, выражающая детерминированный закон эволюции величины X (тренд), $\eta(k)$ – случайная величина, характеризующая отклонение фактического значения показателя от его тренда (здесь и далее будем считать, что $\eta(k)$ – некоррелированная случайная величина с нулевым математическим ожиданием).

Анализ свойств $\eta(k)$ может быть положен в основу задачи классификации клиентов по стадиям взаимоотношений. Идея классификации состоит в группировке объектов анализа (клиентов) по мере однородности анализируемых параметров (по среднеквадратичному отклонению $\eta(k)$, отнесённому к среднему значению $G(k)$, выраженному в процентах) [1]. Введём величину

$$v = \frac{\sigma_\eta}{G} \times 100, \quad (3)$$

где σ_{η} - среднеквадратичное отклонение величины $\eta(k)$, $\bar{G} = \sum_{k=1}^J G(k) / J$ - среднее значение тренда $G(k)$. В случае, если $G(k) = const$, то v есть коэффициент вариации величины X_k . Определив значение v для каждого клиента, его можно отнести к тому или иному классу стадии взаимоотношений. Так, например, при значении $v \in (0,20)$ клиента можно отнести к классу «очень стабильный», при значении $v \in (20,50)$ клиент относится к классу «средне стабильный», при $v \in (50,100)$ - «нестабильный», а при $v > 100$ - «разовый»; границы классов выбираются исходя из специфики деятельности компании. Собственно, для целей функциональности управления взаимоотношениями с клиентами важно не столько значение величины v сколько направление её изменения (то есть «улучшаются» или «ухудшаются» отношения с клиентом).

Основной проблемой определения величины v для каждого клиента является выбор функции $G(k)$ и выбор величин n и J таким образом, чтобы модель задачи была максимально репрезентативной.

Постановка задачи и метод решения

Рассмотрим формулу (2). Выберем в качестве модели тренда $G(k)$ полиномиальную модель

$$G(k) = \sum_{j=0}^m a_j \cdot k^j. \quad (4)$$

Формула (2) может быть записана в обобщённом виде

$$X = Ka + \eta, \quad (5)$$

где X - вектор значений X_i , K - матрица размерности $J \times m$, a - вектор коэффициентов полинома a_j , η - вектор случайных величин $\eta(k)$.

Решение уравнения (5) методом наименьших квадратов относительно вектора a имеет вид¹:

$$\hat{a} = (K^T R^{-1} K)^{-1} K^T R^{-1} X, \quad (6)$$

где \hat{a} - оценка вектора a , $R = M[\eta \eta^T]$ - ковариационная матрица (M - оператор математического ожидания). В случае независимых одинаково распределённых $\eta(k)$ матрица R имеет диагональный вид, так что $R_{kk} = \sigma_{\eta}^2$, где σ_{η} - среднеквадратичное отклонение величины $\eta(k)$. При такой статистической равнозначности $\eta(k)$ формула (6) будет иметь вид

$$\hat{a} = (K^T K)^{-1} K^T X, \quad (7)$$

¹ Крамер Г. Математические методы статистики. - М.: Мир. 1975.

а, наряду с оценкой вектора a , может быть получена и оценка σ_η в виде формулы

$$\hat{\sigma}_\eta^2 = \frac{1}{J - m} (X - K\hat{a})^T (X - K\hat{a}). \quad (8)$$

Если через $\Delta a = \|a - \hat{a}\|$ обозначить погрешность решения задачи (5) методом (7), то соответствующая ей дисперсионная матрица будет иметь вид

$$D_a = M[\Delta[\Delta a^T J] = \left(\frac{1}{\hat{\sigma}_\eta^2} K^T K\right)^{-1}$$

а оценка этой матрицы -

$$\hat{D}_a = \left(\frac{1}{\hat{\sigma}_\eta^2} K^T K\right)^{-1}. \quad (9)$$

Определение функции $G(k)$. Проблема выбора модели тренда $G(k)$ сводится в данном случае к определению степени полинома m . В настоящей работе предлагается способ выбора величины m на основе вероятностной оценки значимости коэффициентов полинома в формуле (4). Пусть \hat{a}_j - оценка соответствующего коэффициент полинома (4), $\hat{D}_a(j, j)$ - коэффициент матрицы (9), лежащий на пересечении j строки и j столбца (т.е. на диагонали). Поскольку распределение вероятности величин \hat{a}_j , найденных по формуле (7), будет близко к нормальному распределению независимо от распределения величин $\eta(k)$ [6], то вероятность того, что нулевые значения величины a_j не входят в область их вероятных значений может быть выражена формулой

$$P(a_j \neq 0) = \int_0^{2|\hat{a}_j|} f_{\hat{a}_j}(\tau) d\tau$$

где $f_{\hat{a}_j}(\tau)$ - функция плотности нормального распределения со средним \hat{a}_j и дисперсией $\hat{D}_a(j, j)$. Решение о том, что a_j отлично от 0 принимается в том случае, если значение $P(a_j \neq 0)$ превышает некоторый порог U_a . Степень полинома (4) определяется, таким образом, максимальным j , для которого $P(a_j \neq 0)$ больше заданного вероятностного порога U_a .

Определение длины периода n . Выбор числа дней n в периоде, за который будут суммироваться данные при анализе, предлагается делать на основе оценки величины $\hat{\sigma}_\eta$ из формулы (8). Рассмотрим величину x_i - значение выбранного показателя в день с номером i . Суть проблемы в

том, что в задачах экономического анализа, как правило, значительную часть выборки x_i составляют нулевые значения (например, продажи какому-либо клиенту могут иметь место лишь 2-3 раза в месяц). Именно поэтому исходные данные x_i принято преобразовывать к некоторым «укрупнённым» данным X_k согласно формуле (1), и встаёт вопрос о выборе длины n такого «укрупнённого» периода таким образом, чтобы среди X_k , по возможности, не встречалось бы нулевых значений (считается, что $x_i \geq 0$).

Пусть $G(k)$ - модель тренда величины X_k , а $\bar{z}_k^{n-m}(G(k))$ - минимальное значение этого тренда на интервале $k = \overline{1, J}$. Пусть σ_η - оценка среднеквадратичного отклонения случайной составляющей модели, полученная по формуле (8). Тогда вероятность того, что найденное Z больше нуля может быть оценена формулой

$$P(Z > 0) = \int_0^\infty f_z(\tau\tau) d, \quad (10)$$

где $f_z(\tau\tau)$ - функция плотности нормального распределения со средним Z и дисперсией σ_η^2 . Выбор n , таким образом, сводится к перебору его возможных значений в порядке возрастания, решения для них задачи (10) и выбора такого n , начиная с которого величина $P(Z > 0)$ станет превышать некоторый порог U_z . При этом задача об определении n должна решаться вместе с задачей об определении степени полинома $G(k)$.

Определение числа периодов J . Выбор числа периодов, по данным которых будет решаться задача (3) должен осуществляться из следующих соображений. С одной стороны, J должно быть достаточным для «сглаживания» единичных «случайных выбросов» величины X_k . С другой - J не должно быть слишком большим, чтобы существенное изменение величины v не было чересчур «сглажено» ретроспективными данными X_k и могло быть обнаружено в самом начале этого процесса.

Пусть ΔX - величина «случайного выброса» X_k , так что $\Delta X = G(k) - X_k$. В том случае, если величина ΔX имеет нормальное распределение со среднеквадратичным отклонением σ_η , можно утверждать, что $|\Delta X| < 2\sigma_\eta$ с вероятностью, приблизительно 0,95. Для практического использования представляется целесообразным выбрать такое J , начиная с которого в случае единичного случайного выброса $|\Delta X| = 2\sigma_\eta$ величина v изменялась бы не более чем на U_v . Конкретное значение U_v

выбирается исходя из границ v для тех или иных классов стадии взаимоотношений (скажем, если границы классов выбраны как 30%, 60% и 90%, то U_v можно принять равным половине интервала - 15%). Вычисление зависимости величины $\Delta v = |v - v_{\Delta X}|$ от J при $|\Delta X| = 2\sigma_\eta$, где $v_{\Delta X}$ - значение величины v , в случае, если в данных имеется случайный выброс ΔX , производится путём моделирования задачи (3) на имеющемся наборе данных X_k .

Результаты численного моделирования

Численное моделирование рассматриваемой задачи классификации клиентов по стадиям взаимоотношений проводилось на реальных данных о годовой динамике продаж в крупной компании, занимающейся торговлей автозапчастями.

На рис. 1 показаны результаты решения задачи для трех разных клиентов (три колонки графиков). На графиках 1а показано значение выручки от продаж товаров по дням в течение года и значение функции тренда $G(k)$, определенной как полином второй степени. На графиках 1б показан результат определения степени полинома (4). Так, видно, что для первого клиента (рис. 1б₁) величина коэффициента a_2 с высокой вероятностью отлична от нуля, следовательно, функцию тренда $G(k)$ оценки продаж этого клиента целесообразно представить полиномом второй степени; для второго и третьего клиента (рис. 1б₂, 1б₃) вероятность того, что коэффициент a_2 отличен от нуля, невелика, а значит, функцию тренда $G(k)$ этих клиентов достаточно моделировать полиномом первой степени (из практического опыта моделирования величину вероятностного порога U_a рекомендуется принимать не менее 0,95).

На графиках 1с показан результат решения задачи выбора числа дней в периоде, за который должны суммироваться данные при анализе. Так, видно, что для первого и второго клиентов (рис. 1с₁, 1с₂) достаточно суммировать 12-13 значений, для третьего (рис. 1с₃) клиента - не менее 25-27. На графиках 1д показаны результаты моделирования задачи (3) для вычисления зависимости величины $\Delta v = |v - v_{\Delta X}|$ от J при $|\Delta X| = 2\sigma_\eta$. Видно, что в случае, если $U_v = 15$, для первого и третьего клиента решение задачи классификации по стадиям взаимоотношений может быть корректно выполнено на данных 4-5 периодов (рис. 1д₁, 1д₃); для второго клиента (рис. 1д₂) таких периодов должно быть не менее 7.

Наконец, на графиках 1е показан результат решения собственно задачи классификации клиентов по стадиям взаимоотношений с определенными согласно предлагаемой методике параметрами. Границы классов принимаются равными 0-30% (постоянный покупатель первого типа - более стабильный), 30-60% (постоянный покупатель второго типа - менее стабильный), более 60% (нестабильный покупатель). Видно, что

первый клиент (рис. 1e₁) в течение всего года находится в пределах класса «постоянный покупатель второго типа», второй клиент (рис. 1e₂) от стадии «нестабильный покупатель» постепенно переходит к пограничному состоянию между двумя классами «постоянный покупатель», третий клиент (рис. 1e₃) в течение года улучшил свою стадию от класса «постоянный покупатель второго типа» до класса «постоянный покупатель первого типа».

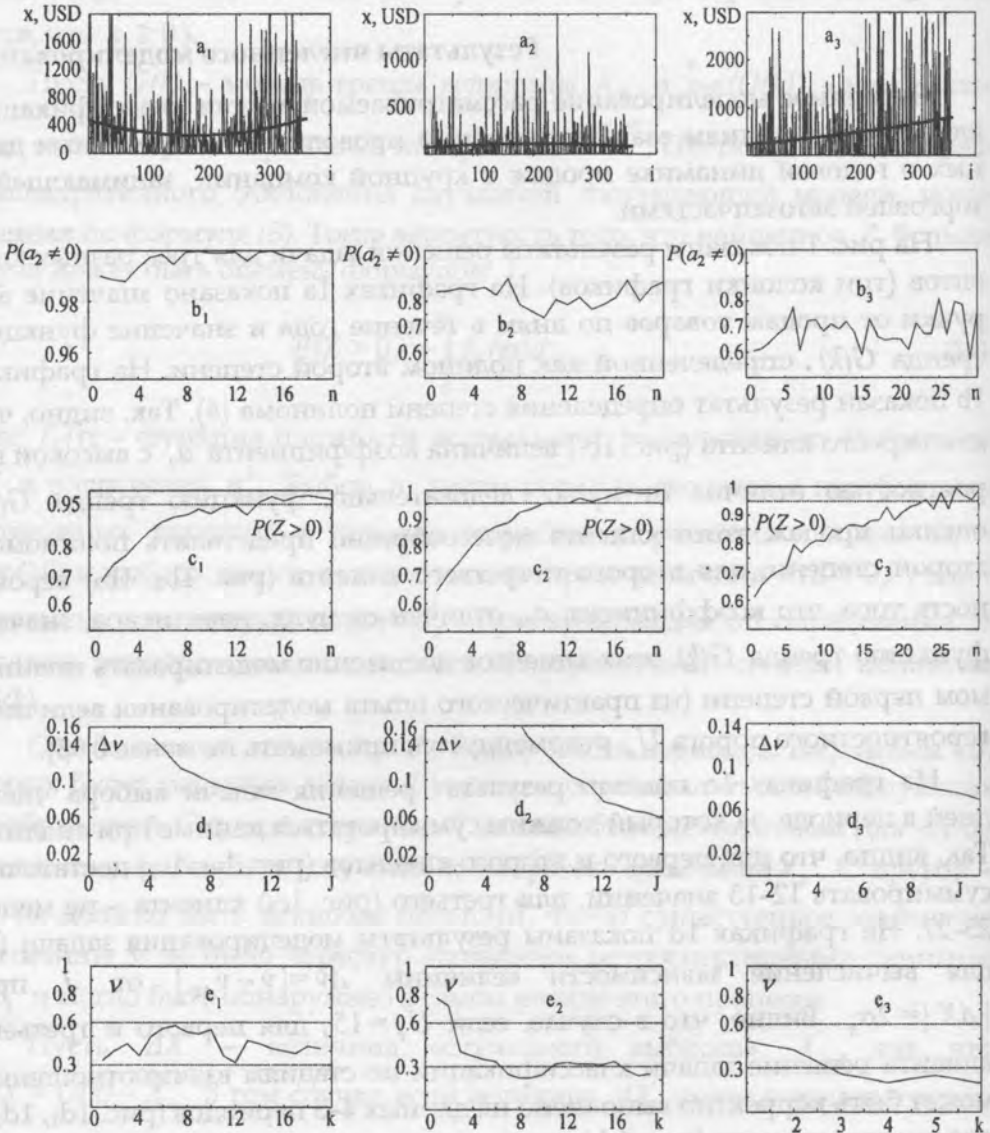


Рис. 1. Результаты моделирования задачи классификации клиентов по стадиям взаимоотношений

В настоящее время предлагаемая в работе методика решения задачи классификации клиентов по стадиям взаимоотношений адаптирована к данным корпоративной информационной системы «1С: Управление торговлей 8» и «1С: Управление производственным предприятием 8» и реализована в виде обработки платформы «1С: 8», основное назначение которой – предоставить менеджеру рекомендации по выбору значений параметров задачи (величины периода анализа, числа периодов и границ классов), наиболее подходящих к использованию в программных модулях указанных типовых конфигураций системы. Представленные в работе результаты демонстрируют конструктивность и непротиворечивость предлагаемой методики. Работа ориентирована на расширение функциональности современных корпоративных информационных систем.