

## МЕТОДЫ ПРОГНОЗИРОВАНИЯ ПЛАТЕЖЕСПОСОБНОСТИ ЗАЁМЩИКОВ

Владимир Николаевич КЛЯЧКИН<sup>а\*</sup>, Юлия Сергеевна ШУНИНА<sup>б</sup>

<sup>а</sup> доктор технических наук, профессор кафедры прикладной математики и информатики, Ульяновский государственный технический университет, Ульяновск, Российская Федерация  
v\_kl@mail.ru

<sup>б</sup> аспирантка кафедры прикладной математики и информатики, Ульяновский государственный технический университет, Ульяновск, Российская Федерация  
ydoncova@yandex.ru

\* Ответственный автор

### История статьи:

Принята 09.07.2015  
Одобрена 28.07.2015

УДК 336.77

JEL: C02, C53, G21

### Аннотация

**Предмет.** В статье рассматривается процесс прогнозирования платежеспособности заемщиков кредитных организаций. Для формирования грамотной стратегии при работе с различными видами задолженностей по кредиту актуальной задачей является разработка более точных методов прогнозирования платежеспособности заемщиков на этапе погашения кредита.

**Цели.** Разработка методов, адекватно описывающих динамику погашения кредита каждого заемщика в зависимости от условий кредита, информации о самом заемщике и его кредитной истории, а также обеспечение достаточной точности прогноза его платежеспособности в следующем периоде.

**Методология.** Предложены методы прогнозирования платежеспособности заемщика на основе марковских цепей 1-го и 2-го порядков, позволяющих учитывать прошлые состояния кредитной истории, совместно с методами машинного обучения для оценки вероятностей переходов из одного состояния кредитного счета в другое с включением факторов, предположительно имеющих влияние на платежеспособность. В рассматриваемом подходе также предлагается провести предварительный анализ данных кредитной истории и восстановить пропущенные значения.

**Результаты.** На основе разработанных методов предложен алгоритм прогнозирования платежеспособности заемщиков. Эффективность подхода показана на примере. Для данных по кредитным счетам реализованы методы прогнозирования, среди которых выявлены наилучшие методы по каждому переходу из одного состояния кредитного счета в другое. При оценке качества расчета найдены матрицы расхождений между реальными и прогнозными значениями.

**Выводы.** Использование предложенных методов повышает результативность прогнозирования платежеспособности для каждого перехода из одного состояния кредитного счета в другое и позволяет кредитной организации заранее принять меры по снижению риска аномальных ситуаций.

### Ключевые слова:

кредитование, задолженность, возврат кредита, марковская цепь, машинное обучение

© Издательский дом ФИНАНСЫ и КРЕДИТ, 2015

### Постановка задачи

Имеется множество заемщиков банка, которым одобрили кредит и которые погашают его в течение заданного срока  $T$ . Известна кредитная история каждого заемщика  $(I_1, \dots, I_t)$ , представляющая собой последовательность состояний кредитного счета, принимающих значение из конечного множества  $S_1, \dots, S_m$  за определенный период  $t$ . Кредитная история заемщика, как правило, характеризует своевременность платежей. Также известен  $X = (x_1, \dots, x_d)$  –  $d$ -мерный вектор признаков, предположительно влияющих на кредитную историю заемщика.

Задача заключается в разработке методов описания динамики состояния кредитного счета заемщика, способных обеспечить с достаточной точностью прогноз состояния счета на следующем шаге  $I_{t+1}$  на основе информации о прошлых состояниях счета  $I_1, \dots, I_t$  с учетом признаков  $X = (x_1, \dots, x_d)$ .

В области кредитования данная задача решается в рамках поведенческого скоринга и имеет также название behavioral scoring [1–2].

В качестве набора признаков  $x_1, \dots, x_d$ , могут выступать следующие данные:

- данные о заемщиках из анкет, которые заполняются при подаче заявки на кредит (фамилия, имя, отчество, дата рождения,

- образование, семейное положение, ежемесячные доходы и расходы, информация об имеющемся имуществе и т.д.);
- сведения о кредите (год взятия, процентная ставка, сумма, коэффициенты LTV (кредит/заклад), DTI (платеж/доход));
- информация о кредитной истории (просрочил ли клиент первый платеж по кредиту, наихудшее состояние счета в прошлом);
- внешние факторы, связанные с состоянием экономической среды и конъюнктурой;
- внутренние факторы, вызванные ошибочными действиями самого банка.

Что касается состояний кредитного счета  $S_1, \dots, S_m$ , то Центральный банк РФ<sup>1</sup> выделяет как для физических лиц, так и для субъектов малого и среднего предпринимательства следующие состояния просроченной задолженности по ссудам:

- без задолженности;
- задолженность от 1 до 30 календарных дней;
- задолженность от 31 до 90 календарных дней;
- задолженность от 91 до 180 календарных дней;
- задолженность свыше 180 календарных дней.

По законодательству РФ каждая кредитная организация должна взаимодействовать с кредитным бюро, отправляя данные как о физических, так и о юридических заемщиках. Это, как правило, информация о заемщике и о том, как он выполняет свои обязательства по кредиту. Согласно данным кредитных бюро<sup>2</sup>, существуют следующие состояния кредитного счета  $S_1, \dots, S_m$ :

- 0 – новый счет, оценка невозможна;
- X – нет информации;
- 1 – оплата без просрочек;
- A – просрочка от 1 до 29 дней;
- 2 – просрочка от 30 до 59 дней;

- 3 – просрочка от 60 до 89 дней;
- 4 – просрочка от 90 до 119 дней;
- 5 – просрочка более 120 дней;
- 7 – регулярные консолидированные платежи;
- 8 – погашение по кредиту с использованием залога;
- 9 – безнадежный долг (передано на взыскание, пропущенный платеж).

### Решение задачи прогнозирования платежеспособности: обзор методов и моделей

Для решения поставленной задачи может быть использована теория марковских цепей, которая была изучена в работах по исследованию динамики структуры кредитного портфеля [3–4]. При этом основная проблема заключается в оценке вероятностей – переходных из одного состояния кредитного счета в другое.

В работе [3] для прогнозирования состояний задолженностей по кредитам авторы предлагают разделить исходную совокупность данных на несколько групп, в которых различаются некоторые вероятности переходов, а затем применить марковскую цепь 1-го порядка, где в качестве оценок переходных вероятностей используют байесовские оценки.

Однако в этом случае нет возможности учета различных факторов, которые могут иметь влияние на оценку платежеспособности заемщика (анкетные данные по кредиту, коэффициенты LTV или DTI и др.).

В работе [4] исследуется динамика структуры кредитного портфеля на основе описания изменения состояния отдельного кредита как марковской цепи 1-го порядка с конечным числом состояний задолженности. С помощью марковской цепи и построения доверительных интервалов для оценок матрицы переходных вероятностей прогнозируется структура портфеля. В качестве оценок переходных вероятностей состояний кредита  $p_{ij}(t)$  принимаются наблюдаемые частоты

переходов:

$$p_{ij}(t) = \frac{f_{ij}(t)}{f_i(t-1)},$$

1 Положение Банка России от 26.03.2004 № 254-П «О порядке формирования кредитными организациями резервов на возможные потери по ссудам, по ссудной и приравненной к ней задолженности» (в ред. от 04.12.2009).

2 URL: <https://www.tcsbank.ru/tournament>.

где  $f_{ij}(t)$  – количество кредитов, перешедших из состояния  $S_i$  в состояние  $S_j$  на шаге  $t$ ;

$f_i(t-1)$  – количество кредитов, находящихся в состоянии  $S_i$  в момент  $t-1$ .

Однако в данном случае задача ориентирована на прогноз структуры портфеля кредитов в целом, то есть нельзя проанализировать платежеспособность отдельно взятого заемщика. Кроме того, также нет возможности учета различных факторов.

Основной недостаток названных работ может быть исправлен путем применения полиномиальной логистической модели [5–7]. Тем не менее она не обладает явным преимуществом перед другими многочисленными классами технологий: нейронными сетями, байесовскими классификаторами, дискриминантным анализом, методом опорных векторов и т.д. [8].

### Подготовка исходных данных

При анализе кредитной истории заемщиков могут возникнуть проблемы, связанные с отсутствующей или пропущенной информацией, которая может привести в итоге к неадекватным результатам. В этом случае используют различные методы восстановления данных [9–10].

Период кредитной истории берется, как правило, равным 1–2 годам: при меньшем сроке можно недооценить вероятность задолженности, а за больший период могут произойти перемены социально-экономического уровня, в результате которых характеристики новых заемщиков будут резко выделяться на фоне старых.

При задании вектора признаков  $X = (x_1, \dots, x_d)$  на результативность прогнозирования могут оказывать влияние наличие корреляционной связи между ними, их значимость, дискретность и непрерывность [11].

Чтобы проверить точность и адекватность методов прогнозирования состояния счета, исходная выборка разбивается на обучающую и контрольную части.

Обучающая выборка используется для построения методов прогнозирования и представляет собой по каждому заемщику набор признаков  $x_1, \dots, x_d$  и кредитную историю  $I_1, \dots, I_{t-1}$ .

Контрольная выборка используется для оценки точности построенных методов путем сравнения реального и прогнозируемого состояний счета  $I_t$  для каждого заемщика.

### Предлагаемые методы прогнозирования состояния счета заемщика

Кредитная история отдельно взятого заемщика представляет собой последовательность дискретно распределенных случайных величин  $I_1, \dots, I_t$ , каждая из которых принимает значение из некоторого фиксированного множества  $S_1, \dots, S_m$ . При этом вероятность того, что кредитный счет заемщика находится в данном состоянии в данный момент, можно определить на основе предшествовавших состояний. Таким образом, последовательность  $I_1, \dots, I_t$  можно назвать цепью Маркова [12–13].

Марковская цепь позволяет на основе оценки матрицы переходных вероятностей спрогнозировать платежеспособность заемщика, то есть состояние его кредитного счета в следующем периоде.

Порядок марковской цепи определяется количеством предшествовавших состояний, которые влияли на текущее состояние счета. Поскольку результаты прогнозирования на основе марковских цепей 3-го и более высоких порядков, как показали расчеты, не улучшают результатов прогнозирования, полученных с использованием марковских цепей меньших порядков, ограничимся рассмотрением марковских цепей 1-го и 2-го порядков.

Введем следующие обозначения:

$v_i(t)$  – вероятность того, что счет окажется в состоянии  $S_i$  в момент  $t$ , где  $i = 1, \dots, m$ , причем  $0 \leq v_i(t) \leq 1$ ,  $v_1(t) + \dots + v_m(t) = 1$ ;

$v(t)$  – вектор вероятностей состояний счета в момент  $t$ ,  $v(t) = [v_1(t), \dots, v_m(t)]$ ;

$P(t)$  – матрица переходных вероятностей марковской цепи 1-го порядка;

$p_{ij}(t)$  – вероятность перехода счета из состояния  $S_i$  в момент  $t$  в состояние  $S_j$  за один шаг:  
 $p_{ij}(t) \equiv P(I_{t+1} = S_j \mid I_t = S_i)$ ;

$\varphi_k(t)$  – вероятность того, что счет окажется в состоянии  $S_k$  в момент  $t$ , если предыдущими

состояниями были  $S_i$  и  $S_j$ , где  $i, j = 1, \dots, m$ , причем  $0 \leq \varphi_k(t) \leq 1$ ,  $\varphi_1(t) + \dots + \varphi_\eta(t) = 1$  (где  $\eta$  – количество состояний с учетом двух предыдущих состояний);

$\varphi(t)$  – вектор вероятностей состояний счета в момент  $t$ :

$$\begin{aligned} \varphi(t) &= [\varphi_1(t), \dots, \varphi_\eta(t)] = \\ &= [(v_1(t-1), v_1(t)), \dots, (v_1(t-1), v_m(t)), \\ &(v_m(t-1), v_1(t)), \dots, (v_m(t-1), v_m(t))]; \end{aligned}$$

$\Theta(t)$  – матрица переходных вероятностей марковской цепи 2-го порядка;

$p_{ijk}(t)$  – вероятность перехода счета в состояние  $S_k$ , если предыдущими состояниями были  $S_i$  в момент  $t-1$  и  $S_j$  в момент  $t$ ,  $p_{ijk}(t) \equiv P(I_{t+1} = S_k \parallel I_t = S_j, I_{t-1} = S_i)$ .

Предположим, что вероятности перехода не зависят от номера шага, т  $p_{ij}(t) \equiv p_{ij}$ . Тогда

динамику состояния счета отдельно взятого заемщика можно описать с помощью марковской цепи 1-го и 2-го порядков соответственно:

$$v_j(t+1) = \sum_{i=1}^m p_{ij} v_i(t),$$

$$\varphi_k(t+1) = \sum_{i=1}^m \sum_{j=1}^m p_{ijk} (v_i(t-1), v_j(t)).$$

Запишем эти модели для векторов вероятностей состояний счета отдельного заемщика:

$$v(t+1) = v(t)P(t, t+1),$$

$$\varphi(t+1) = \varphi(t)\Theta(t, t+1).$$

Предполагается, что матрицы переходных вероятностей  $P$  и  $\Theta$  точно не известны, и необходимо найти их оценки, чтобы спрогнозировать состояние счета заемщика на следующем шаге.

Воспользуемся различными методами машинного обучения для построения зависимости между оценками переходных вероятностей и факторами, предположительно влияющими на платежеспособность заемщиков  $X = (x_1, \dots, x_d)$ .

Поскольку состояние счета  $S_i$  может перейти в одно из  $m$  возможных состояний, то каждую строку матрицы переходных можно описать с помощью полиномиальной модели, которую представим в виде набора бинарных моделей.

Тогда будем рассматривать отдельные переходы из одного состояния в другое ( $S_i \rightarrow S_j$  для марковской цепи 1-го порядка и переходы  $S_i \rightarrow S_j \rightarrow S_k$  – для марковской цепи 2-го порядка).

Таким образом, в качестве зависимой переменной будут выступать оценки переходных вероятностей для марковских цепей 1-го и 2-го порядков соответственно:

$$\hat{p}_{ij} = \begin{cases} 1, & \text{если был совершен переход } S_i \rightarrow S_j, \\ 0, & \text{если иначе} \end{cases}$$

$$\hat{p}_{ijk} = \begin{cases} 1, & \text{если был совершен переход } S_i \rightarrow S_j \rightarrow S_k, \\ 0, & \text{если иначе} \end{cases}$$

Построим зависимости между оценками переходных вероятностей для марковской цепи 1-го и 2-го порядков, а также факторами, предположительно влияющими на платежеспособность заемщиков  $X = (x_1, \dots, x_d)$  на основе методов машинного обучения:

- на основе дискриминантного анализа (ДА)<sup>3</sup>:

$$\hat{p}_{ij} = q_0 + q_1 x_1 + \dots + q_d x_d,$$

где  $q_0, \dots, q_d$  – параметры регрессии;

- на основе логистической регрессии (ЛР) [5]:

$$\hat{p}_{ij} = \frac{\exp(q_0 + q_1 x_1 + \dots + q_d x_d)}{1 + \exp(q_0 + q_1 x_1 + \dots + q_d x_d)},$$

$$\hat{p}_{ijk} = \frac{\exp(q_0 + q_1 x_1 + \dots + q_d x_d)}{1 + \exp(q_0 + q_1 x_1 + \dots + q_d x_d)};$$

- на основе нейронной сети (НС) [15]:

$$\Omega = \sum_{r=1}^d q_r X_r + \alpha,$$

$$\hat{p}_{ij} = F(\Omega), \hat{p}_{ijk} = F(\Omega),$$

где  $\Omega$  – результат суммирования;

$q_r$  – вес  $r$ -го синапса;

$X_r$  – входной сигнал;

$\alpha$  – значение смещения;

$\hat{p}_{ij}$  – выходной сигнал;

$F$  – функция активации;

- на основе метода опорных векторов (МОВ) [16]:

$$\hat{p}_{ij} = F(X) = \text{sign}(\langle \mu, X \rangle + \beta),$$

<sup>3</sup> Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы. М.: Финансы и статистика, 2003. 352 с.

$$\hat{p}_{ijk} = F(X) = \text{sign}(\langle \mu, X \rangle + \beta),$$

где  $\langle \cdot, \cdot \rangle$  – скалярное произведение;

$\mu$  – нормальный вектор к разделяющей гиперплоскости;

$X$  – вектор-столбец признаков;

$\beta$  – вспомогательный параметр;

- на основе наивного байесовского классификатора (НБК) [17–18]:

$$\hat{p}_{ij} = P(S_i \rightarrow S_j | X = x) = \frac{P(S_i \rightarrow S_j) \prod_{r=1}^d P(X_r = x_r | S_i \rightarrow S_j)}{\sum_{(i \rightarrow j)'} P(S_i \rightarrow S_j) \prod_{r=1}^d P(X_r = x_r | (S_i \rightarrow S_j)')};$$

$$\hat{p}_{ijk} = P(S_i \rightarrow S_j \rightarrow S_k | X = x) = \frac{P(S_i \rightarrow S_j \rightarrow S_k) \prod_{r=1}^d P(X_r = x_r | S_i \rightarrow S_j \rightarrow S_k)}{\sum_{(i \rightarrow j)'} P(S_i \rightarrow S_j \rightarrow S_k) \prod_{r=1}^d P(X_r = x_r | (S_i \rightarrow S_j \rightarrow S_k)')}.$$

- на основе деревьев решений (ДР)<sup>4</sup>: для каждого перехода  $S_i \rightarrow S_j$  и  $S_i \rightarrow S_j \rightarrow S_k$  строится свое дерево решений.

Предлагаемый алгоритм прогнозирования платежеспособности заемщиков:

1. Обработка исходных данных и приведение их к определенному виду (табл. 1);
2. Восстановление пропусков в кредитной истории ( $I = M$ ) при помощи моделирования с использованием условного распределения по присутствующим параметрам [20]. Чтобы сделать это, необходимо найти общую для всех заемщиков матрицу переходных вероятностей (из одного состояния кредитного счета в другие) определенного вида (табл. 2).

При этом  $p_{ij}$  находится по формуле

$$p_{ij}(t) = \frac{f_{ij}(t)}{f_i(t-1)},$$

где  $f_{ij}(t)$  – количество счетов, перешедших из состояния  $S_i$  в  $S_j$  на шаге  $t$ ;

$f_i(t-1)$  – количество счетов, находящихся в состоянии  $S_i$  в момент  $t-1$ .

При обнаружении пропуска в качестве  $S_i$  используется предыдущее состояние кредитной истории, а восстановленное состояние можно найти с помощью построенной матрицы переходных вероятностей, которая будет иметь максимальное значение вероятности перехода из состояния  $S_i$ .

3. Формирование обучающей и контрольной выборок. В случае кредитной истории за 12 мес. обучающая выборка будет представлять собой по каждому заемщику набор признаков  $x_1, \dots, x_d$  и кредитную историю за 11 мес., а контрольная выборка – состояние кредитной истории за 12-й месяц по каждому заемщику.

4. Построение разработанных моделей описания динамики погашения кредита на обучающей выборке с помощью марковских цепей и нахождение оценок вероятностей с помощью методов классификации для всех переходов  $S_i \rightarrow S_j$  и  $S_i \rightarrow S_j \rightarrow S_k$ . Тогда для марковских цепей получатся свои матрицы переходных вероятностей (табл. 3, 4).

5. Построение прогноза по каждому заемщику на следующем шаге  $I_{t+1}$ , то есть определение состояния кредитного счета на 12-й месяц. Для этого в случае марковской цепи 1-го порядка необходимо определить для каждого счета текущее состояние  $I_t$  и подставить значения факторов  $x_1, \dots, x_d$  в модели. Тогда результат прогнозирования каждого счета будет представлять собой распределение вероятностей перехода из текущего состояния  $I_t$  в следующее  $I_{t+1}$ :

$I_t$	$I_{t+1}$				
	$S_1$	$S_2$	$S_3$	...	$S_m$
$S_i$	$\hat{p}_{i1}$	$\hat{p}_{i2}$	$\hat{p}_{i3}$	...	$\hat{p}_{im}$

В случае марковской цепи 2-го порядка необходимо определить для каждого счета текущее состояние  $I_t$  и предыдущее состояние  $I_{t-1}$  и подставить значения факторов  $x_1, \dots, x_d$  в модели.

Тогда результат прогнозирования каждого счета

4 Якутов А.И. Применение деревьев решений для моделирования кредитоспособности клиентов коммерческого банка // Искусственный интеллект. 2008. № 4. С. 208–213.

будет представлять собой следующее распределение вероятностей перехода:

$I_{t-1}, I_t$		$I_{t+1}$				
		$S_1$	$S_2$	$S_3$	...	$S_m$
$S_i$	$S_j$	$\hat{P}_{ij1}$	$\hat{P}_{ij2}$	$\hat{P}_{ij3}$	...	$\hat{P}_{ijm}$

1. Определение прогнозируемого состояния  $I_{t+1}$  с помощью максимальной вероятности перехода.
2. Анализ результатов прогнозирования и оценка качества построенных методов с помощью нахождения матрицы расхождений (табл. 5). Ее строки представляют собой реальное состояние счета заемщика на следующем шаге  $I_{t+1}$ , а столбцы – прогнозируемое состояние на следующем шаге  $I_{t+1}$ . Таким образом, на диагонали находится общее количество верно прогнозируемых состояний  $s_{aa}$ , а в остальных ячейках – количество ошибок первого и второго рода  $s_{ab}$ , где  $a = 1, \dots, m, b = 1, \dots, m$ .

Процент верных прогнозов  $\varepsilon$ , усредненный по всем имеющимся состояниям, может быть найден по формуле

$$\varepsilon = \frac{\sum_{a=1}^m \frac{s_{aa}}{\sum_{b=1}^m s_{ab}}}{m} \cdot 100\%.$$

3. Выбор для каждого перехода метода, наилучшим образом прогнозирующего будущее состояние счета (с наибольшей долей правильных прогнозов), и формирование для конкретных исходных данных матрицы переходных вероятностей, наилучшим образом прогнозирующей состояние счета по всем переходам.

**Пример.** В качестве примера были рассмотрены данные по 8 500 кредитным счетам под залог недвижимости. Кредитная история каждого заемщика представлена ежемесячными состояниями кредитного счета в период с июня 2011 г. по июнь 2012 г. В качестве состояний кредитного счета представлены следующие состояния задолженности:

- С – без задолженности;
- 30 – 0–29 дней просрочки;

- 60 – 30–59 дней просрочки;
- 90 – 60–89 дней просрочки;
- 120 – 90–119 дней просрочки;
- F – 120–149 дней просрочки (переход права собственности на заложенное имущество на кредитору);
- R – более 149 дней просрочки (продажа заложенного имущества кредитором для покрытия задолженности).

В качестве факторов, описывающих кредитный счет, представлена информация об условиях кредита (дата получения кредита, его сумма, коэффициенты LTV, ДТІ и т.д.).

На основе данной выборки был составлен прогноз платежеспособности заемщиков. В качестве среды программирования использована программа MATLAB.

1. Были определены допустимые (+) и недопустимые (–) переходы из одного состояния счета в другие (табл. 6).
2. Кредитная история по всем счетам была разделена на обучающую выборку, представляющую собой последовательность ежемесячных состояний счетов в период с июня 2011 г. по май 2012 г., и на контрольную выборку – состояния счетов в июне 2012 г.
3. Был проведен процесс восстановления пропущенных данных по кредитной истории.
4. В качестве факторов, влияющих на платежеспособность заемщиков, были приняты текущий коэффициент LTV и коэффициент ДТІ.
5. Для каждого перехода из одного состояния в другие по всем счетам были реализованы предложенные методы платежеспособности на основе марковской цепи (МЦ) 1-го и 2-го порядков.

В табл. 7 представлены проценты верных прогнозов по всем возможным переходам из одного состояния  $S_i$  в другие и средний процент по всем переходам для каждого метода, а также максимальные значения для каждого перехода.

Результаты исследования показали, что нет единого метода, наилучшим образом прогнозирующего все переходы. Оптимальным вариантом является выбор наилучшего метода или одного из лучших способов для каждого перехода,

что в среднем улучшает результаты отдельных методов примерно на 7–20%. Таким образом, наилучший результат достигается при использовании нескольких методов прогнозирования (табл. 8).

**Выводы**

В работе предложены методы описания динамики состояния кредитного счета отдельного заемщика. Их особенностью является возможность учета

прошлых состояний счета, а также факторов, предположительно влияющих на данное состояние.

На основе разработанных методов предложен алгоритм прогнозирования состояния кредитного счета в следующем месяце, что позволяет работникам кредитной или коллекторской организации оценить платежеспособность заемщиков и принять соответствующие меры к имеющим просроченную задолженность.

Таблица 1

Исходные данные для алгоритма

№ счета заемщика	Признаки $x_1, \dots, x_d$	Кредитная история $I_1, \dots, I_t$
1	$X_{11}, \dots, X_{1d}$	$I_{11}, \dots, I_{1t}$
2	$X_{21}, \dots, X_{2d}$	$I_{21}, \dots, I_{2t}$
3	$X_{31}, \dots, X_{3d}$	$I_{31}, \dots, I_{3t}$
...	...	...
$n$	$X_{n1}, \dots, X_{nd}$	$I_{n1}, \dots, I_{nt}$

Таблица 2

Общая матрица переходных вероятностей

$S_i$	$S_j$				
	$S_1$	$S_2$	$S_3$	...	$S_m$
$S_1$	$p_{11}$	$p_{12}$	$p_{13}$	...	$p_{1m}$
$S_2$	$p_{21}$	$p_{22}$	$p_{23}$	...	$p_{2m}$
...	...	...	...	...	...
$S_m$	$p_{m1}$	$p_{m2}$	$p_{m3}$	...	$p_{mm}$

Таблица 3

Матрица переходных вероятностей для марковской цепи 1-го порядка

$S_i$	$S_j$				
	$S_1$	$S_2$	$S_3$	...	$S_m$
$S_1$	$\hat{p}_{m11}$	$\hat{p}_{m12}$	$\hat{p}_{m13}$	...	$\hat{p}_{m1m}$
$S_2$	$\hat{p}_{m21}$	$\hat{p}_{m22}$	$\hat{p}_{m23}$	...	$\hat{p}_{m2m}$
...	...	...	...	...	...
$S_m$	$\hat{p}_{m1}$	$\hat{p}_{m2}$	$\hat{p}_{m3}$	...	$\hat{p}_{mm}$

Таблица 4

Матрица переходных вероятностей для марковской цепи 2-го порядка

$S_i \rightarrow S_j$	$S_k$				
	$S_1$	$S_2$	$S_3$	...	$S_m$
$S_1$	$\hat{p}_{111}$	$\hat{p}_{112}$	$\hat{p}_{113}$	...	$\hat{p}_{11m}$
$S_1$	$\hat{p}_{121}$	$\hat{p}_{122}$	$\hat{p}_{123}$	...	$\hat{p}_{12m}$
...	...	...	...	...	...
$S_m$	$\hat{p}_{mm1}$	$\hat{p}_{mm2}$	$\hat{p}_{mm3}$	...	$\hat{p}_{mmm}$

Таблица 5  
Матрица расхождений

Реально	Прогноз				
	$S_1$	$S_2$	$S_3$	...	$S_m$
$S_1$	$S_{11}$	$S_{12}$	$S_{13}$	...	$S_{1m}$
$S_2$	$S_{21}$	$S_{22}$	$S_{23}$	...	$S_{2m}$
...	...	...	...	...	...
$S_m$	$S_{m1}$	$S_{m2}$	$S_{m3}$	...	$S_{mm}$

Таблица 6  
Значения переходов

Реально	Прогноз						
	$C$	30	60	90	120	$F$	$R$
$C$	+	+	–	–	–	–	–
30	+	+	+	–	–	–	–
60	+	+	+	+	–	–	–
90	+	+	+	+	+	–	–
120	+	+	+	+	+	+	–
$F$	+	+	+	+	+	+	+
$R$	+	+	+	+	+	+	+

Таблица 7  
Результаты прогнозирования

$S_i$	МЦ 1-го порядка						МЦ 2-го порядка					Максимум
	ЛР	ДА	НБК	ДР	МОВ	НС	ЛР	ДА	НБК	ДР	НС	
$C$	99,5	99,5	99,5	99,5	99,5	99,5	99,7	99,7	99,7	99,5	99,7	99,7
30	49,6	49,6	49,6	49,6	49,6	49,6	38,1	38,1	38,1	44,2	35,4	49,6
60	33,3	33,3	33,3	33,3	40	40	40	33,3	26,7	20	33,3	40
90	0	0	33,3	50	0	25	50	50	50	50	50	50
120	75	75	33,3	41,7	75	58,3	41,7	50	41,7	16,7	50	75
$F$	100	100	100	100	100	90,4	98,6	98,6	98,6	97,3	90,4	100
$R$	55,6	55,6	66,7	55,6	55,6	88,9	66,7	66,7	55,6	33,3	0	88,9
Среднее	59	59	59,4	61,4	59,9	64,5	62,1	62,3	58,6	51,6	51,3	71,9

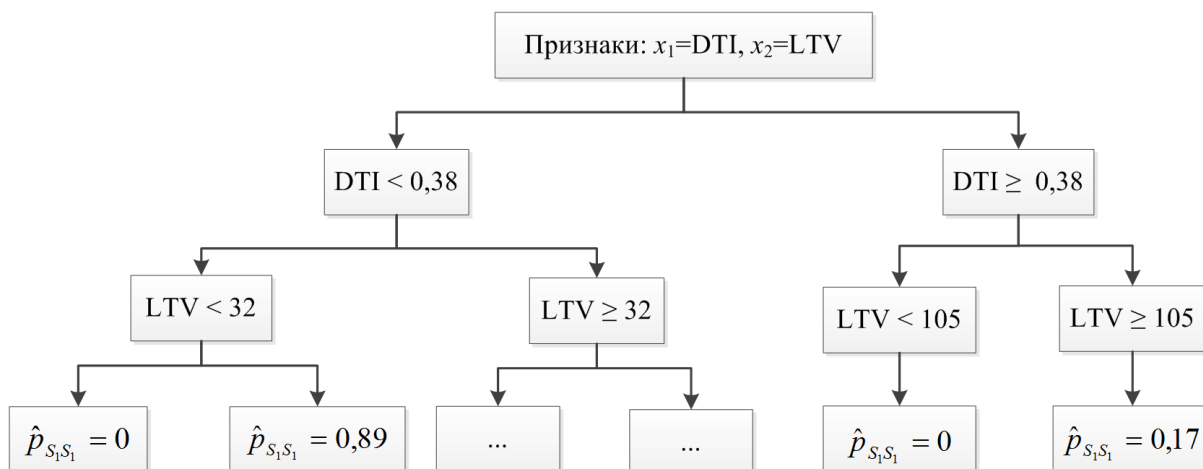
Таблица 8  
Наиболее результативные методы прогнозирования

Вид МЦ	$I_t$	$I_{t+1}$					
		$C$	30	60	90	120	$R$
МЦ 2-го порядка	$C$	Логистическая регрессия: $\hat{p}_{ijk} = \frac{\exp(q_0 + q_1 x_1 + \dots + q_d x_d)}{1 + \exp(q_0 + q_1 x_1 + \dots + q_d x_d)}$					
МЦ 1-го порядка	30	Логистическая регрессия: $\hat{p}_{ij} = \frac{\exp(q_0 + q_1 x_1 + \dots + q_i x_d)}{1 + \exp(q_0 + q_1 x_1 + \dots + q_i x_d)}$					
	60	Метод опорных векторов: $\hat{p}_{ij} = F(X) = \text{sign}(\langle \mu, X \rangle + \beta)$					
	90	Деревья решений					
	120	Логистическая регрессия: $\hat{p}_{ij} = \frac{\exp(q_0 + q_1 x_1 + \dots + q_i x_d)}{1 + \exp(q_0 + q_1 x_1 + \dots + q_i x_d)}$					
	$R$	Нейронная сеть: $\hat{p}_{ij} = F(\Omega), \Omega = \sum_{r=1}^n q_r X_r + \alpha$					



Рисунок 1

Дерево решений для перехода  $S_1 \rightarrow S_1$  с учетом признаков  $x_1$  и  $x_2$



### Список литературы

1. *Thomas L.C.* A survey of credit and behavioral scoring: forecasting financial risk of lending to consumers // *International Journal of Forecasting*. 2000. Vol. 16. Iss. 2. P. 149–172.
2. *Thomas L.C., Ho J., Scherer W.T.* Time will tell: behavioral scoring and the dynamics of consumer credit assessment // *IMA Journal of Management Mathematics*. 2001. Vol. 12. Iss. 1. P. 89–103.
3. *Grimshaw S., Alexander W.* Markov chain models for delinquency: Transition matrix estimation and forecasting // *Applied Stochastic Models in Business and Industry*. 2011. № 27. Vol. 27. Iss. 3. P. 267–279.
4. *Тимофеева Г.А., Тимофеев Н.А.* Прогнозирование составляющих кредитного портфеля на основе модели марковской цепи // *Автоматика и телемеханика*. 2012. № 4. С. 47–65.
5. *Diggle P.J., Liang K.Y., Zeger S.L.* *Analysis of Longitudinal Data*. Oxford: Oxford University Press, 2002.
6. *Клячкин В.Н., Донцова (Шунина) Ю.С.* Сравнительный анализ точности нелинейных моделей при прогнозировании состояния системы на основе марковской цепи // *Известия Самарского научного центра Российской академии наук*. 2013. Т. 15. № 4. С. 924–927.
7. *Чижова А.С.* Эконометрическая модель оценки матриц вероятностей перехода кредитных рейтингов // *Прикладная эконометрика*. 2007. № 3. С. 11–26.
8. *Мерков А.Б.* *Распознавание образов: введение в методы статистического обучения*. М.: URSS, 2010. 254 с.
9. *Айвазян С.А., Енюков И.С., Мешалкин Л.Д.* *Прикладная статистика. Основы моделирования и первичная обработка данных*. М.: Финансы и статистика, 1983. 471 с.
10. *Литтл Р. Дж. А., Рубин Д.Б.* *Статистический анализ данных с пропусками / пер. с англ.* М.: Финансы и статистика, 1990. 336 с.
11. *Гринь Н.В.* Методологические аспекты построения скоринговых моделей // *Экономика, моделирование, прогнозирование: сб. науч. тр. Вып. 6*. Минск: НИЭИ Минэкономики РБ, 2012. С. 174–180.
12. *Pfeifer P.E. Robert L.* Modeling customer relationships as Markov Chains. *Journal of Interactive Marketing*. 2000. № 14. Iss. 2. P. 43–55.

13. *Соколов Г.А., Чистякова Н.А.* Теория вероятностей. Управляемые цепи Маркова в экономике. М.: ФИЗМАТЛИТ, 2005. 248 с.
14. *Уоссермен Ф.* Нейрокомпьютерная техника: теория и практика / пер. Ю.А. Зуева, В.А. Точенова. М.: Мир, 1992. 184 с.
15. *Ванник В.Н.* Восстановление зависимостей по эмпирическим данным. М.: Наука, 1979. 448 с.
16. *Вагин В.Н., Головина Е.Ю., Загорянская А.А., Фомина М.В.* Достоверный и правдоподобный вывод в интеллектуальных системах. М.: ФИЗМАТЛИТ, 2004. 704 с.
17. *Thomas L.C.* Credit Scoring and its Applications: SIAM Monographs on Mathematical Modeling and Computation / L.C. Thomas, D.V. Edelman, J.N. Crook. SIAM: Philadelphia, USA. 2002. 248 p.
18. *Алексеева В.А., Донцова Ю.С., Клячкин В.Н.* Восстановление пропущенных наблюдений при классификации объектов // Известия Самарского научного центра Российской академии наук. 2014. Т. 16. № 6. С. 357–359.

## CUSTOMER SOLVENCY PREDICTION METHODS

Vladimir N. KLYACHKIN<sup>a,\*</sup>, Yuliya S. SHUNINA<sup>b</sup>

<sup>a</sup> Ulyanovsk State Technical University, Ulyanovsk, Russian Federation  
ydoncova@yandex.ru

<sup>b</sup> Ulyanovsk State Technical University, Ulyanovsk, Russian Federation  
v\_kl@mail.ru

\* Corresponding author

### Article history:

Received 9 July 2015  
Accepted 28 July 2015

**JEL classification:** C02, C53, G21

**Keywords:** lending, debt, loan repayment, Markov chain, machine learning

### Abstract

**Importance** The article reviews the process of forecasting the solvency of the credit institution's customers. To outline an appropriate strategy for handling various loan balances, it is reasonable to devise more accurate methods for forecasting the solvency of borrowers at the loan repayment stage.

**Objectives** The research pursues setting up methods, which would adequately reflect each borrower's tendency of repaying the loan depending on the loan terms, borrower's profile and credit history, and ensuring sufficient accuracy of the borrower's solvency forecast for the following period.

**Methods** We proposed methods for predicting the borrower's solvency through Markov chains of the 1<sup>st</sup> and 2<sup>nd</sup> order that allowed considering the previous status of the credit history, and machine learning methods to evaluate probabilities of the credit account changes, including factors that could possibly influence the ultimate solvency. The approach implies a preliminary analysis of the credit history data and recovers missing indicators.

**Results** Based on the methods, we propose our own algorithm for forecasting the borrowers' solvency. The efficiency of the approach is illustrated with an example. As for data of credit accounts, we implemented prediction methods and identified the best ones per each change in the status of the credit account. When evaluating the quality of estimates, we found matrices of differences between the real and projected indicators.

**Conclusions and Relevance** The proposed methods will improve the solvency forecasts for each change in the credit account's status and enable the credit institution to take relevant measures to reduce the anomaly risk

© Publishing house FINANCE and CREDIT, 2015

## References

1. Thomas L.C. A Survey of Credit and Behavioural Scoring: Forecasting Financial Risk of Lending to Consumers. *International Journal of Forecasting*, 2000, vol. 16, iss. 2, pp. 149–172.
2. Thomas L.C., Ho J., Scherer W.T. Time will Tell: Behavioral Scoring and the Dynamics of Consumer Credit Assessment. *IMA Journal of Management Mathematics*, 2001, vol. 12, iss. 1, pp. 89–103.
3. Grimshaw S., Alexander W. Markov Chain Models for Delinquency: Transition Matrix Estimation and Forecasting. *Applied Stochastic Models in Business and Industry*, 2011, vol. 27, iss. 3, pp. 267–279.
4. Timofeeva G.A., Timofeev N.A. Prognozirovanie sostavlyayushchikh kreditnogo portfelya na osnove modeli markovskoi tsepi [Forecasting credit portfolio components with a Markov chain model]. *Avtomatika i telemekhanika = Automation and Remote Control*, 2012, no. 4, pp. 47–65.
5. Diggle P.J., Liang K.Y., Zeger S.L. *Analysis of Longitudinal Data*. Oxford, Oxford University Press, 2002.
6. Klyachkin V.N., Dontsova Yu.S. Sravnitel'nyi analiz tochnosti nelineynykh modelei pri prognozirovanii sostoyaniya sistemy na osnove markovskoi tsepi [A comparative analysis of the accuracy of nonlinear models in forecasting the condition of system on the basis of the Markov chain]. *Izvestiya Samarskogo nauchnogo tsentra Rossiiskoi akademii nauk = Proceedings of Samara Scientific Center of RAS*, 2013, vol. 15, no. 4, pp. 924–927.

7. Chizhova A.S. Ekonometricheskaya model' otsenki matrits veroyatnostei perekhoda kreditnykh reitingov [An econometric model for evaluating the probability matrices of changes in credit ratings]. *Prikladnaya ekonometrika = Applied Econometrics*, 2007, no. 3, pp. 11–26.
8. Merkov A.B. *Raspoznavanie obrazov: vvedenie v metody statisticheskogo obucheniya* [Pattern recognition: an introduction into statistical learning methods]. Moscow, URSS Publ., 2010, 254 p.
9. Aivazyan S.A., Enyukov I.S., Meshalkin L.D. *Prikladnaya statistika. Osnovy modelirovaniya i pervichnaya obrabotka dannykh* [Applied statistics. Fundamentals of modeling and primary processing of data]. Moscow, Finansy i statistika Publ., 1983, 471 p.
10. Little R.J.A., Rubin D.B. *Statisticheskii analiz dannykh s propuskami* [Statistical Analysis with Missing Data]. Moscow, Finansy i statistika Publ., 1990, 336 p.
11. Grin' N.V. *Metodologicheskie aspekty postroeniya skoringovykh modelei. V kn.: Ekonomika, modelirovanie, prognozirovanie. Vypusk 6* [Methodological aspects of setting scoring models. In: Economics, modeling, forecasting. Issue 6]. Minsk, R&D Institute of Ministry of Economy of Republic of Belarus Publ., 2012, pp. 174–180.
12. Pfeifer P.E., Carraway R.L. Modeling Customer Relationships as Markov Chains. *Journal of Interactive Marketing*, 2000, vol. 14, iss. 2, pp. 43–55.
13. Sokolov G.A., Chistyakova N.A. *Teoriya veroyatnostei. Upravlyaemye tsepi Markova v ekonomike* [Probability theory. Controlled Markov chains in economy]. Moscow, FIZMATLIT Publ., 2005, 248 p.
14. Wasserman P. *Neirokomp'yuternaya tekhnika: teoriya i praktika* [Neural Computing: Theory and Practice]. Moscow, Mir Publ., 1992, 184 p.
15. Vapnik V.N. *Vosstanovlenie zavisimostei po empiricheskim dannym* [Restoring correlations through empirical data]. Moscow, Nauka Publ., 1979, 448 p.
16. Vagin V.N., Golovina E.Yu., Zagoryanskaya A.A., Fomina M.V. *Dostoverniy i pravdopodobnyi vyvod v intellektual'nykh sistemakh* [Credible and plausible inference in intelligent systems]. Moscow, FIZMATLIT Publ., 2004, 704 p.
17. Thomas L.C., Edelman D.B., Crook J.N. *Credit Scoring and its Applications: SIAM Monographs on Mathematical Modeling and Computation*. SIAM, Philadelphia, USA, 2002, 248 p.
18. Alekseeva V.A., Dontsova Yu.S., Klyachkin V.N. Vosstanovlenie propushchennykh nablyudenii pri klassifikatsii ob"ektov [Missing observations recovery during objects classification]. *Izvestiya Samarskogo nauchnogo tsentra Rossiiskoi akademii nauk = Proceedings of Samara Scientific Center of Russian Academy of Sciences*, 2014, vol. 16, no. 6, pp. 357–359.