

**АДАПТАЦИЯ МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ КРЕДИТОСПОСОБНОСТИ  
С УЧЕТОМ ВНОВЬ ПОСТУПАЮЩЕЙ ИНФОРМАЦИИ О КЛИЕНТАХ****Виктор Ростиславович КРАШЕНИННИКОВ<sup>а</sup>, Владимир Николаевич КЛЯЧКИН<sup>б</sup>,  
Юлия Сергеевна ШУНИНА<sup>с</sup>\***<sup>а</sup> доктор технических наук, профессор, заведующий кафедрой прикладной математики и информатики,  
Ульяновский государственный технический университет,  
Ульяновск, Российская Федерация  
kvrulstu@mail.ru<sup>б</sup> доктор технических наук, профессор кафедры прикладной математики и информатики,  
Ульяновский государственный технический университет,  
Ульяновск, Российская Федерация  
v\_kl@mail.ru<sup>с</sup> кандидат технических наук, инженер-программист Ульяновского конструкторского бюро приборостроения,  
Ульяновск, Российская Федерация  
ydoncova@yandex.ru

\* Ответственный автор

**История статьи:**

Получена 20.12.2016

Получена в доработанном  
виде 17.03.2017

Одобрена 24.03.2017

Доступна онлайн 15.06.2017

УДК 336.77

JEL: C02, C53, G21

<https://doi.org/10.24891/fa.10.6.663>**Ключевые слова:** возврат  
кредита, метод  
прогнозирования,  
машинное обучение,  
псевдоградиентная  
процедура,  
структура модели**Аннотация****Тема.** Для прогнозирования возврата кредитов заемщиками используются методы машинного обучения. В последнее время широко распространено применение различных видов агрегации этих методов. Актуальной задачей является обновление структуры агрегированного метода прогнозирования по истечении некоторого времени для адаптации к особенностям характеристик вновь поступивших клиентов, а также получения более точного прогноза.**Цели.** Разработка способа обновления структуры названного метода для повышения точности прогнозирования.**Методология.** Использовались методы машинного обучения с агрегированием различных классификаторов на основе нейронной сети, логистической регрессии, дискриминантного анализа, наивного байесовского классификатора, метода опорных векторов и др. Для корректировки параметров моделей используется псевдоградиентная процедура. Качество полученной структуры модели с обновленными параметрами оценивается по среднему квадрату ошибки на контрольной выборке.**Результаты.** Разработан способ обновления структуры модели прогнозирования возврата кредитов, эффективность которого подтверждена практическими испытаниями на данных российских заемщиков.**Выводы.** Использование псевдоградиентной процедуры корректировки параметров выбранной модели позволяет получить приемлемую точность прогнозирования на некоторый период. Однако при значительных изменениях кредитной ситуации эта модель не в состоянии обеспечить достаточную точность прогноза ни при каких значениях своих параметров. Поэтому время от времени нужно менять структуру модели, чтобы она в новых условиях обеспечивала достаточную точность прогнозирования и получение более обоснованных решений о выдаче кредита новым клиентам.**Применение.** Предложенный способ может быть применен в автоматизированных системах для поддержки принятия решений о выдаче кредитов в банковской сфере.

© Издательский дом ФИНАНСЫ и КРЕДИТ, 2016

**Постановка задачи**Рассматривается задача прогнозирования  
возврата кредитов в банковской сфере.  
Имеется множество заемщиков  $Z_1 \dots Z_m$ ,которым банк выдал кредиты. По каждому  
заемщику  $Z$  известен набор анкетных данных  
(возраст, пол, семейное положение, размер и  
срок кредита и пр.), характеризующийся  
 $n$ -мерным вектором  $X_r = (x_{r1} \dots x_{rn})$ .

По прошествии некоторого времени проясняется информация о возврате кредита: заемщик вернул его или имеется задолженность. Таким образом, заемщики оказываются разделенными на два класса:  $C_1$ , состоящий из выплативших кредит, и  $C_2$  – из имеющих задолженность.

По этой информации требуется построить классификатор, с помощью которого можно было бы с наибольшей точностью отнести новых клиентов к одному из двух классов в случае выдачи им кредитов. При этом следует учесть возможность того, что кредитная ситуация и характеристики новых клиентов со временем могут изменяться.

Таким образом, искомым классификатор по вектору признаков  $X_s = (x_{s1} \dots x_{sn})$  нового клиента  $K_s$  должен отнести его к классу  $C_1$  (выплатит кредит) или к  $C_2$  (будет иметь задолженность).

### Подходы к решению задачи

Введем функцию, определенную на составе заемщиков [1]:

$$P(Z_r) = \begin{cases} 1, & \text{если заемщик } Z_r \text{ вернул кредит,} \\ 0, & \text{если заемщик } Z_r \text{ имеет задолженность.} \end{cases} \quad (1)$$

По существу,  $P(X_r)$  есть апостериорная (по факту) вероятность возврата кредита заемщиком  $Z_r$ . Построим оценку  $\hat{P}(X_s)$  вероятности возврата кредита новым клиентом  $K_s$  в виде функции его вектора параметров  $X_s$ .

Естественно, построение этой оценки проводится по данным о заемщиках (1). Разобьем эти сведения на обучающую и контрольную *выборки*. Возьмем некоторую модель, то есть функцию  $\hat{P}(X_r)$ , параметры которой определим так, чтобы на обучающей выборке средний квадрат

$$\sigma_1^2 = \frac{1}{l} \sum_{r=1}^l [P(X_r) - \hat{P}(X_s)]^2$$

ошибки оценки был минимален. Качество полученной оценки определяется средним квадратом ошибки  $\sigma_2^2$  на контрольной выборке.

Классификация, то есть отнесение клиента  $K_s$  к тому или иному классу, будет производиться исходя из значения  $\hat{P}(X_r)$ . Но, естественно, классификация будет тем лучше, чем точнее  $\hat{P}(X_r)$  оценивает  $P(X_r)$ . Таким образом, сначала требуется построить функцию (оценку)  $\hat{P}(X_r)$  с минимальным средним квадратом отклонения от  $P(X_r)$ .

Существует множество способов решения поставленной задачи. Методы кластерного анализа могут быть применимы для задачи кластеризации, то есть обучения без учителя, когда изначально в исходных данных класс кредитоспособности не известен и требуется искать зависимость между объектами, чтобы сгруппировать их в кластеры.

Однако используемые авторами исходные данные содержат информацию о клиенте, а также класс кредитоспособности, то есть налицо задача обучения с учителем.

Классический метод такой учебы – дискриминантный анализ, но в методах машинного обучения этот вид анализа рассматривается как один из нескольких.

Действительно, этот метод не всегда дает лучший результат. Поэтому в данном случае применяются методы машинного обучения, которые в процессе построения учатся верно сопоставлять набору признаков о клиенте определенный класс кредитоспособности.

После построения методов, имея информацию о новых клиентах, можно определить их класс, но вместо этого мы определяем вероятность возврата кредита, чтобы можно было не просто выдать или не выдать его, а предложить на разных условиях.

Известен целый ряд классификаторов на основах [2–8]<sup>1</sup>:

- нейронной сети (НС);

<sup>1</sup> Шунина Ю.С., Алексеева В.А., Клячкин В.Н. Прогнозирование кредитоспособности клиентов банка на основе методов машинного обучения // *Финансы и кредит*. 2015. № 27. С. 2–12; Вагин В.Н., Головина Е.Ю., Загорянская А.А., Фомина М.В. Достоверный и правдоподобный вывод в интеллектуальных системах / под ред. В.Н. Вагина, Д.А. Поспелова. М.: Физматлит, 2004. 704 с.

- логистической регрессии (ЛР);
- дискриминантного анализа (ДА);
- наивного байесовского классификатора (НБК);
- метода опорных векторов (МОВ);
- дерева решений (ДР);
- бэггинга деревьев решений (БДР) [9].

В качестве *нейронной сети* будем использовать трехслойную сеть прямого распространения, архитектура которой представлена на *рис. 1*.

Узел (нейрон) сети состоит из трех логических блоков: входы, функция преобразования и выход. Блок функции преобразования включает в свою очередь синаптические веса  $q_1 \dots q_n$ , суммирующую функцию и одну из функций нелинейного преобразования (гиперболический тангенс или логарифм) (*рис. 2*).

Количество нейронов во входном слое определяется количеством признаков заемщика, а в скрытом слое – по теореме Колмогорова.

Выходной слой представлен одним нейроном, содержащим вероятность возврата кредита.

Для обучения нейронной сети используется самый популярный метод обучения многослойной сети – метод обратного распространения ошибки, в процессе которого сеть учится верно классифицировать заемщиков по анкетным данным на группы кредитоспособности.

*Логистическая регрессия* основана на предположении о том, что вероятность кредитоспособного клиента  $P\{C_1 | X\}$  равна

$$f(z), z = q_0 + q_1 x_1 + \dots + q_n x_n,$$

где  $f(z)$  – логистическая функция:

$$f(z) = 1 / (1 + e^{-z}). \quad (2)$$

*Метод опорных векторов* заключается в переводе исходных векторов в пространство более высокой размерности, а также поиске оптимальной гиперплоскости, разделяющей классы клиентов наилучшим образом. Классифицирующая функция имеет следующий вид:

$$f(X) = \text{sign}(\langle a, X \rangle + \beta),$$

где  $\langle a, X \rangle$  – скалярное произведение;

$a$  – нормальный вектор к разделяющей гиперплоскости;

$\beta$  – параметр, который определяется кратчайшим расстоянием между гиперплоскостью и началом координат.

*Дискриминантный анализ* основан на применении формулы Байеса:

$$\hat{P}(X) = P(C_1 | X) = \frac{P(C_1)G_1(X)}{P(C_1)G_1(X) + P(C_2)G_2(X)},$$

где  $P(C_1 | X)$  – условная вероятность возврата кредита клиентом при наборе признаков  $X$ ;

$P(C_1)$ ,  $P(C_2)$  – априорные вероятности принадлежности заемщиков к классам  $C_1$  и  $C_2$  соответственно, а также предположение о нормальном распределении входных данных с математическими ожиданиями  $\psi_1$  и  $\psi_2$  и ковариационными матрицами  $\Sigma_1$  и  $\Sigma_2$ :

$$G_1(X) = \frac{1}{(2\pi|\Sigma_1|)^{1/2}} \exp[-1/2(X - \psi_1)^T \Sigma_1^{-1}(X - \psi_1)],$$

$$G_2(X) = \frac{1}{(2\pi|\Sigma_2|)^{1/2}} \exp[-1/2(X - \psi_2)^T \Sigma_2^{-1}(X - \psi_2)].$$

*Наивный байесовский классификатор* основан также на применении формулы Байеса и предположении об условной независимости входных данных:

$$\begin{aligned} \hat{P}(X) &= P(C_1 | X) = \\ &= \frac{P(C_1) \prod_{u=1}^n P(x_u | C_1)}{P(C_1) \prod_{u=1}^n P(x_u | C_1) + P(C_2) \prod_{u=1}^n P(x_u | C_2)}, \end{aligned}$$

где  $P(x_u | C_1)$ ,  $P(x_u | C_2)$  – вероятности наличия признаков  $x_1 \dots x_n$  в классах  $C_1$  и  $C_2$  соответственно.

В качестве *дерева решений* используется дерево классификации (рис. 3).

Вероятность возврата кредита есть вероятность принадлежности заемщика классу  $C_1$ , которая определяется как доля заемщиков данного класса в конечном узле, содержащем общее количество заемщиков для двух классов.

Для построения оптимального дерева решений применяется индекс Джини:

$$Gini(L) = 1 - \sum_{m=1}^2 d_m^2,$$

где  $L$  – узел дерева;

$d_m$  – доля заемщиков класса  $m$  в конечном узле.

*Бэггинг деревьев решений* основан на применении нескольких деревьев решений на разных частях обучающей выборки. Общий результат метода находится с помощью процедуры голосования, то есть клиенту присваивается тот класс, которому отдает предпочтение большинство деревьев.

Каждый из перечисленных методов имеет свои преимущества и недостатки, поэтому сложно говорить о рейтинговой значимости. Например, логистическая регрессия учитывает ограничения на значения вероятности, которые не могут выходить за рамки 0 и 1, а также построение модели на основе не измененной начальной выборки и реальной пропорции между количеством кредитоспособных и некредитоспособных клиентов. Однако модель чувствительна к корреляции между факторами.

Дискриминантный анализ удобен в использовании при наличии нескольких классов кредитоспособности заемщиков.

Нейронные сети наилучшим образом проявили себя при прогнозировании возврата кредитов юридическими лицами.

К преимуществам деревьев решений относится быстрая обработка больших объемов данных, работа с пропущенными, числовыми и нечисловыми типами данных, а также отсутствие ограничений на коррелируемость между входными переменными.

К преимуществам байесовского подхода также относится возможность вывода по отсутствующим данным, однако необходимо обязательное выполнение предположения о независимости данных о заемщике.

В статье предлагается совместное использование некоторых методов путем агрегирования результатов по среднему значению, что позволяет компенсировать недостатки одних моделей при помощи других и повысить точность прогнозирования [10–12].

*Агрегированный классификатор* по среднему значению имеет следующий вид:

$$P_{AK\text{cp}}^{\wedge}(X_r) = \frac{\sum_{K=1}^H \hat{P}_K(X_r)}{H},$$

где  $\hat{P}_K(X_r)$  – вероятность возврата кредита  $r$ -м клиентом, найденная с помощью  $K$ -го метода, причем  $r = 1 \dots l$  и  $K = 1 \dots H$ , где  $H$  – количество методов в наборе.

### Обновление модели классификатора

В связи с изменением внешней среды компании, использующие в своей практике методы прогнозирования, нуждаются в периодическом обновлении последних. Поскольку новые данные являются самыми значимыми, особенно для краткосрочного прогноза, они вводятся в обучающую выборку. На ее основе можно заново построить классификатор, но это может оказаться трудоемким процессом.

После построения классификатора возникает необходимость в учете новых клиентов, данные о

которых со временем могут существенно меняться в зависимости от социально-экономической ситуации, условий кредитования, а также характеристик самих клиентов.

Предлагается обновление параметров моделей на основе численного метода, заключающегося в корректировке коэффициентов моделей, входящих в состав агрегированного метода, с применением псевдоградиента [13–15]<sup>2</sup>.

Эта процедура позволяет учитывать сведения о новых клиентах без пересчета данных о старых клиентах. Другие методы адаптации математических моделей к вновь поступающей информации рассмотрены в трудах российских и зарубежных специалистов [16, 17]<sup>3</sup>.

Оптимизация модели состоит в нахождении вектора ее параметров  $\bar{q}$ , минимизирующих средний квадрат ошибок:

$$\Omega(\bar{q}) = \sum_{r=1}^l [P(X_r) - \hat{P}(X_r, \bar{q})]^2, \quad (3)$$

где  $\hat{P}(X_r, \bar{q})$  – оценка вероятности возврата кредита, зависящая от признаков  $X$  клиента и вектора параметров модели  $\bar{q}$ .

Суть псевдоградиентной процедуры заключается в постоянной корректировке вектора параметров модели при поступлении информации о каждом новом клиенте на основе следующей формулы:

$$\bar{q}_{i+1} = \bar{q}_i - v_i \nabla J(\bar{q}_i),$$

где  $\bar{q}_{i+1}$  – следующее за  $\bar{q}_i$  приближение оптимального вектора  $\bar{q}$ ;

$v_i$  – коэффициенты, влияющие на величину шага;

$\nabla J(\bar{q}_i) = \nabla [P(Y_i) - \hat{P}(X_i, \bar{q}_i)]^2$  – градиент отдельного слагаемого из формулы (3), который является псевдоградиентом функции  $\Omega(\bar{q})$ .

<sup>2</sup> Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов. М.: Финансы и статистика, 2003. 416 с.

<sup>3</sup> Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. М.: ЮНИТИ, 1998. 1024 с.

Корректировка параметров модели достигается путем добавления к старому вектору параметров  $\bar{q}_i$  поправки, получаемой в результате умножения числа  $v_i$  на псевдоградиент  $\nabla J(\bar{q}_i)$ .

При этом важно, что эта процедура способна отслеживать меняющуюся ситуацию, так как текущий вектор  $\bar{q}_i$  постоянно на нее реагирует в соответствии с характеристиками новых клиентов.

Однако по истечении некоторого периода структура агрегированного метода прогнозирования может устареть, и применение псевдоградиентной процедуры (которая оптимизирует только параметры модели) может показать далеко не самый лучший результат. В таком случае предлагается произвести обновление структуры агрегированного классификатора.

### Обновление структуры классификатора

Рассмотрим имеющиеся данные по российским заемщикам за 2006–2011 гг.<sup>4</sup> Примем в качестве обучающей выборки данные за 2006–2009 гг. Оценим прогнозирование отдельных и агрегированных методов классификации на данных клиентов за 2010 г. (табл. 1) по среднему квадрату ошибки.

Согласно результатам табл. 1, начальной оптимальной структурой является агрегированный метод классификации.

Пусть имеется множество новых клиентов (2010–2011 гг.), данные о которых поступают последовательно. С помощью псевдоградиентной процедуры обновим параметры модели логистической регрессии. Тогда  $\hat{P}(X_i, \bar{q}_i)$  находится по следующей формуле:

$$\hat{P}(X_i, \bar{q}_i) = f(z(\bar{q}_i)),$$

$$z(\bar{q}_i) = q_{i0} + q_{i1} x_{i1} + \dots + q_{in} x_{in},$$

<sup>4</sup> Данные предоставлены международной компанией по анализу и обработке данных АлгоМост в 2015 г. и находятся в открытом доступе: URL: [http://algomost.com/ru/tasks/uploadfiles/58/train\\_utf\\_noid.sas7bdat](http://algomost.com/ru/tasks/uploadfiles/58/train_utf_noid.sas7bdat)

где  $f(z(\bar{q}_l))$  – логистическая функция:

$$f(z(\bar{q}_l)) = 1/[1 + e^{-z(\bar{q}_l)}].$$

По определению

$$\nabla J(\bar{q}_l) = \frac{d}{d\bar{q}_l} [P(Y_l) - \hat{P}(X_l, \bar{q}_l)]^2,$$

поэтому преобразуем выражение (2) к следующему виду:

$$\begin{aligned} \bar{q}_{l+1} &= \bar{q}_l - v_l \nabla J(\bar{q}_l) = \\ &= \bar{q}_l - v_l \frac{d}{d\bar{q}_l} [P(Y_l) - \hat{P}(X_l, \bar{q}_l)]^2 = \\ &= \bar{q}_l + 2v_l [P(Y_l) - \hat{P}(X_l, \bar{q}_l)] \frac{d\hat{P}(X_l, \bar{q}_l)}{d\bar{q}_l} = \\ &= \bar{q}_l + \eta_l [P(Y_l) - \hat{P}(X_l, \bar{q}_l)] \frac{d\hat{P}(X_l, \bar{q}_l)}{d\bar{q}_l}, \end{aligned}$$

где  $\eta_l = 2v_l$  – новый параметр шага процедуры.

Тогда корректировка коэффициентов логистической регрессии выполняется по формуле

$$\bar{q}_{l+1} = \bar{q}_l + \eta_l [P(Y_l) - \hat{P}(X_l, \bar{q}_l)] \frac{d\hat{P}(X_l, \bar{q}_l)}{d\bar{q}_l},$$

где производные принимают следующий вид:

$$\begin{aligned} \frac{d}{d\bar{q}_l} \hat{P}(X_l, \bar{q}_l) &= \frac{d}{d\bar{q}_l} f(z(\bar{q}_l)) = \\ &= \frac{d}{d\bar{q}_l} \frac{1}{1 + e^{-(q_{l0} + q_{l1}x_{l1} + \dots + q_{ln}x_{ln})}} = \\ &= \frac{e^{-(q_{l0} + q_{l1}x_{l1} + \dots + q_{ln}x_{ln})}}{[1 + e^{-(q_{l0} + q_{l1}x_{l1} + \dots + q_{ln}x_{ln})}]^2} (1, x_{l1} \dots x_{ln}). \end{aligned}$$

Оценим качество методов прогнозирования отдельно для трех различных контрольных выборок (табл. 2).

По результатам рассмотрения приведенных данных можно заметить, что для последней контрольной выборки результат агрегированного метода хуже итогов, полученных на основе нейронной сети и дискриминантного анализа. Это может свидетельствовать о том, что рассмотренная структура агрегированного классификатора устарела, и требуется ее пересмотр.

Добавим в обучающую выборку информацию о новых клиентах, кроме данных третьей контрольной выборки, и оценим качество (табл. 3). При этом агрегированный классификатор резко изменяет структуру.

Таким образом, результат классификации с использованием новой структуры агрегированного метода на 24% превосходит старый вариант. Следовательно, с появлением новых клиентов целесообразно производить обновление не только параметров классификатора, но и его структуру.

## Выводы

В работе предложен способ обновления структуры агрегированного классификатора для улучшения прогнозирования возврата кредитов в банковской сфере. На протяжении краткосрочного периода возможно использование псевдоградиентной процедуры обновления параметров моделей, которая позволяет учитывать информацию о новых клиентах без пересчета моделей на старых клиентах.

Предложенный способ обновления структуры метода обеспечивает адаптацию к меняющимся со временем условиям кредитования, что существенно повышает точность прогнозирования.

**Таблица 1****Средний квадрат ошибок методов прогнозирования на клиентах 2010 г.****Table 1****The average squared error of forecasting techniques in relation to clients, 2010**

Метод	Средний квадрат ошибки
НС	0,249
ДА	0,258
НБК	0,323
МОВ	0,241
ДР	0,427
ЛР	0,239
БДР	0,237
АК*	0,231

\* Здесь АК – метод, агрегирующий результаты логистической регрессии и бэггинга деревьев решений по среднему значению:  $AK = ЛР + БДР$ .

Источник: авторская разработка

Source: Authoring

**Таблица 2****Средний квадрат ошибок методов прогнозирования для трех выборок****Table 2****The average squared error of forecasting techniques for the three samples**

№ выборки	Метод прогнозирования							
	НС	ДА	НБК	МОВ	ДР	ЛР	БДР	АК
1	0,251	0,277	0,357	0,259	0,418	0,263	0,277	0,249
2	0,249	0,257	0,44	0,313	0,445	0,324	0,252	0,249
3	0,256	0,26	0,441	0,305	0,416	0,313	0,31	0,28

Источник: авторская разработка

Source: Authoring

**Таблица 3****Средний квадрат ошибок методов прогнозирования для контрольной выборки № 3****Table 3****The average squared error of forecasting techniques for test sample no. 3**

Метод	Средний квадрат ошибки
НС	0,249
ДА	0,247
НБК	0,278
МОВ	0,242
ДР	0,461
ЛР	0,247
БДР	0,27
АК*	0,226

\* Здесь АК – метод, агрегирующий результаты нейронной сети, наивного байесовского классификатора, метода опорных векторов, дерева решений и бэггинга деревьев решений по среднему значению:

$$AK = НС + НБК + МОВ + ДР + БДР.$$

Источник: авторская разработка

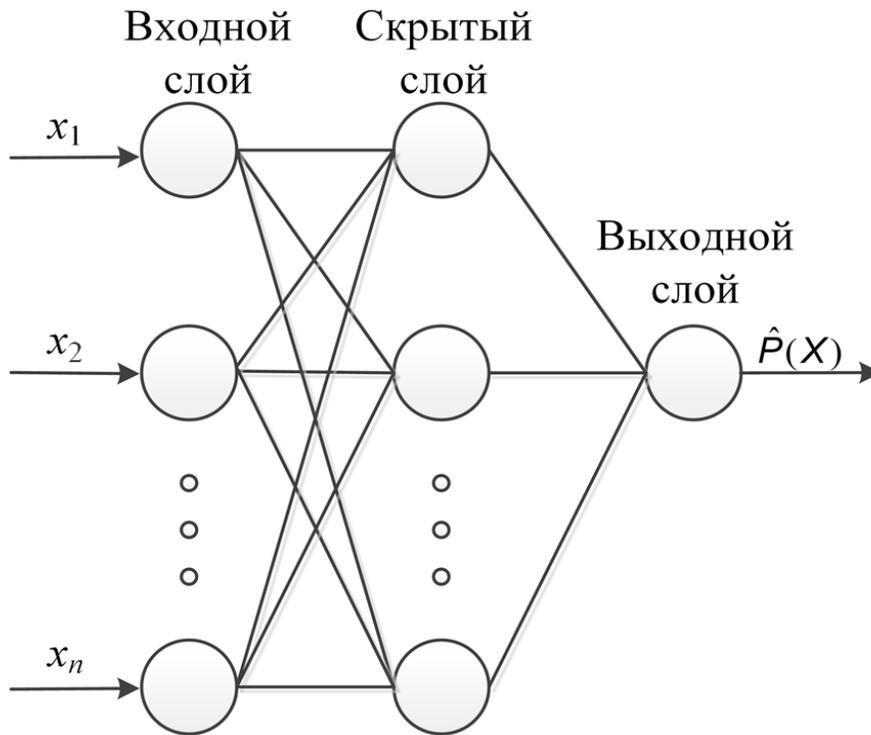
Source: Authoring

*Рисунок 1*

**Архитектура нейронной сети**

*Figure 1*

**Architecture of the neural network**



*Источник:* авторская разработка

*Source:* Authoring



4. *Сорокин А.С.* Построение скоринговых карт с использованием модели логистической регрессии // Интернет-журнал Науковедение. 2014. № 2. URL: <http://naukovedenie.ru/PDF/180EVN214.pdf>
5. *Уоссермен Ф.* Нейрокомпьютерная техника: теория и практика / перевод Ю.А. Зуева, В.А. Точенова. М.: Мир, 1992. 240 с.
6. *Якупов А.И.* Применение деревьев решений для моделирования кредитоспособности клиентов коммерческого банка // Искусственный интеллект. 2008. № 4. С. 208–213.
7. *Мерков А.Б.* Распознавание образов: введение в методы статистического обучения. М.: URSS, 2010. 254 с.
8. *Breiman L.* Random Forests. *Machine Learning*, 2001, vol. 45, iss. 1, pp. 5–32. doi: 10.1023/A:1010933404324
9. *Клячкин В.Н., Шулгина Ю.С.* Система оценки кредитоспособности заемщиков и прогнозирования возврата кредитов // Вестник компьютерных и информационных технологий. 2015. № 11. С. 45–51.
10. *Breiman L.* Bagging predictors. *Machine Learning*, 1996, vol. 24, no. 2, pp. 123–140. doi: 10.1023/A:1018054314350
11. *Ghodselahi A., Amirmadhi A.* Application of Artificial Intelligence Techniques for Credit Risk Evaluation. *International Journal of Modeling and Optimization*, 2011, vol. 1, no. 3, pp. 243–249.
12. *Крашенинников В.Р., Клячкин В.Н., Шулгина Ю.С.* Обновление агрегированных классификаторов на основе псевдоградиентной процедуры // Вестник компьютерных и информационных технологий. 2016. № 10. С. 36–40.
13. *Васильев К.К., Крашенинников В.Р.* Статистический анализ изображений. Ульяновск: УлГТУ, 2014. 214 с.
14. *Montgomery D.C.* A Note on Forecasting with Adaptive Filtering. *Oper. Res. Quart.*, 1977, vol. 28, no. 11, pp. 87–91.
15. *Поляк Б.Т., Цыпкин Я.З.* Псевдоградиентные алгоритмы адаптации и обучения // Автоматика и телемеханика. 1973. № 3. С. 45–68.
16. *Бокс Дж., Дженкинс Г.* Анализ временных рядов. Прогноз и управление. М.: Мир, 1974. 242 с.
17. *Льюис К.Д.* Методы прогнозирования экономических показателей. М.: Финансы и статистика, 1986. 133 с.

### **Информация о конфликте интересов**

Мы, авторы данной статьи, со всей ответственностью заявляем о частичном и полном отсутствии фактического или потенциального конфликта интересов с какой бы то ни было третьей стороной, который может возникнуть вследствие публикации данной статьи. Настоящее заявление относится к проведению научной работы, сбору и обработке информации, написанию и подготовке статьи, принятию решения о публикации рукописи.

**ADAPTATION OF MODELS PREDICTING CUSTOMERS' CREDITWORTHINESS  
IN LINE WITH NEW INCOMING DATA ON CLIENTS****Viktor R. KRASHENINNIKOV<sup>a</sup>, Vladimir N. KLYACHKIN<sup>b</sup>, Yuliya S. SHUNINA<sup>c\*</sup>**<sup>a</sup> Ulyanovsk State Technical University, Ulyanovsk, Russian Federation  
kvrulstu@mail.ru<sup>b</sup> Ulyanovsk State Technical University, Ulyanovsk, Russian Federation  
v\_kl@mail.ru<sup>c</sup> JSC Ulyanovsk Instrument Manufacturing Design Bureau, Ulyanovsk, Russian Federation  
mamaeva\_29\_12@mail.ru

\* Corresponding author

**Article history:**Received 20 December 2016  
Received in revised form  
17 March 2017  
Accepted 24 March 2017  
Available online 15 June 2017**JEL classification:** C02,  
C53, G21<https://doi.org/10.24891/fa.10.6.663>**Keywords:** loan repayment,  
forecasting approach, machine  
learning, pseudo-gradient  
procedure, model structure**Abstract****Importance** Machine learning methods are used to predict repayment of loans by borrowers. It is an acute task to update the structure of the aggregate forecast after some time in order to adapt to new clients' characteristics as well as to provide sufficient prediction accuracy.**Objectives** The aim is to provide the structure updating of the aggregate forecasting method to improve the prediction accuracy.**Methods** In this paper, we used the machine learning methods with different classifiers based on a neural network, logistic regression, discriminant analysis, naive Bayes classifier, Support Vector Machines, etc. To adjust parameters of the models, pseudo-gradient procedure is used. The quality of the model structure obtained with updated parameters is assessed by the average square error in the control sample.**Results** We developed a model structure updating method for forecasting customers' loan repayment, the effectiveness of which has been confirmed by practical tests based on the data according to Russian borrowers.**Conclusions and Relevance** Using the pseudo-gradient procedure of adjusting the parameters of the chosen model provides an accurate prediction for a certain period of time. However, after significant changes in the credit situation, this model is not able to provide a sufficient accuracy of forecast for any values of its parameters. Therefore, from time to time it is needed to change the structure of the model to provide sufficient prediction accuracy to get more justified decisions about granting credits to new customers on the new terms. The proposed method can be applied in automated systems to support decision-making about granting loans in the banking sector.

© Publishing house FINANCE and CREDIT, 2016

**References**

1. Aivazyan S.A., Bukhshtaber V.M., Enyukov I.S. *Prikladnaya statistika: Klassifikatsiya i snizhenie razmernosti* [Applied Statistics: Classification and Dimensionality Reduction]. Moscow, Finansy i Statistika Publ., 1989, 607 p.
2. Vasil'ev N.P., Egorov A.A. [Experience of logistic regression parameters calculation by Newton-Rafson method to estimation resistance to cold of plants]. *Matematicheskaya biologiya i bioinformatika = Mathematical Biology and Bioinformatics*, 2011, vol. 6, no. 2, pp. 190–199. (In Russ.)
3. Bidyuk P.I., Terent'yev A.N., Gasanov A.S. Construction and Methods of Learning of Bayesian Networks. *Cybernetic and System Analysis*, 2005, vol. 41, no. 4, pp. 587–598.
4. Sorokin A.S. [Building of scoring cards using logistic regression]. *Naukovedenie*, 2014, iss. 2. (In Russ.) Available at: <http://naukovedenie.ru/PDF/180EVN214.pdf>

5. Wasserman P. *Neirokomp'yuternaya tekhnika: teoriya i praktika* [Neural Computer Technics: Theory and Practice]. Moscow, Mir Publ., 1992, 240 p.
6. Yakupov A.I. [The decision tree's adaptation for modeling of commercial bank clients solvency]. *Iskusstvennyi intellekt = Artificial Intelligence*, 2008, no. 4, pp. 208–213. (In Russ.)
7. Merkov A.B. *Raspoznavanie obrazov: vvedenie v metody statisticheskogo obucheniya: monografiya* [Image recognition: An introduction into statistical learning methods: a monograph]. Moscow, Editorial URSS Publ., 2010, 254 p. (In Russ.)
8. Breiman L. Random Forests. *Machine Learning*, 2001, vol. 45, iss. 1, pp. 5–32. doi: 10.1023/A:1010933404324
9. Klyachkin V.N., Shunina Yu.S. [System for borrowers' creditworthiness assessment and repayment of loans forecasting]. *Vestnik komp'yuternykh i informatsionnykh tekhnologii = Herald of Computer and Information Technologies. Scientific, Technical and Production Monthly Journal*, 2015, no. 11, pp. 45–51. (In Russ.)
10. Breiman L. Bagging Predictors. *Machine Learning*, 1996, vol. 24, iss. 2, pp. 123–140. doi: 10.1023/A:1018054314350
11. Ghodselahi A., Amirmadhi A. Application of Artificial Intelligence Techniques for Credit Risk Evaluation. *International Journal of Modeling and Optimization*, 2011, vol. 1, iss. 3, pp. 243–249.
12. Krasheninnikov V. R., Klyachkin V. N., Shunina Yu. S. [Pseudo-gradient Procedure as a Method of aggregated Classifiers Updating]. *Vestnik komp'yuternykh i informatsionnykh tekhnologii = Herald of Computer and Information Technologies. Scientific, Technical and Production Monthly Journal*, 2016, no. 10, pp. 36–40. (In Russ.)
13. Vasil'ev K.K., Krasheninnikov V.R. *Statisticheskii analiz izobrazhenii: monografiya* [Statistic analysis of images: a monograph]. Ulyanovsk, Ulyanovsk State Technical University Publ., 2014, 214 p.
14. Montgomery D.C. A Note on Forecasting with Adaptive Filtering. *Operational Research Quarterly*, 1977, vol. 28, no. 11, pp. 87–91.
15. Polyak B.T., Tsytkin Ya.Z. [Pseudogradient adaptation and training algorithms]. *Avtomatika i telemekhanika = Automation and Remote Control*, 1973, no. 3, pp. 45–68. (In Russ.)
16. Box G., Jenkins G.M., Reinsel G. *Analiz vremennykh ryadov. Prognoz i upravlenie* [Time Series Analysis: Forecasting & Control]. Moscow, Mir Publ., 1974, 242 p.
17. Lewis C.D. *Metody prognozirovaniya ekonomicheskikh pokazatelei* [Industrial and Business Forecasting Methods]. Moscow, Finansy i Statistika Publ., 1986, 133 p.

### Conflict-of-interest notification

We, the authors of this article, bindingly and explicitly declare of the partial and total lack of actual or potential conflict of interest with any other third party whatsoever, which may arise as a result of the publication of this article. This statement relates to the study, data collection and interpretation, writing and preparation of the article, and the decision to submit the manuscript for publication.