

**Tregubov Roman Borisovich** – Academy of the Federal Guard Service of the Russian Federation; e-mail: smv\_57@bk.ru; 35, Priborostroitel'naya street, Orel, 302034, Russia; cand. of eng. sc.; member.

**Stremouhov Mihail Vladimirovich** – e-mail: smv\_57@bk.ru; member.

УДК 004.89

**С.С. Алхасов, А.Н. Целых**

### **ОСНОВНЫЕ ПОДХОДЫ К ПОСТРОЕНИЮ ИНФОРМАЦИОННОЙ СИСТЕМЫ ДЛЯ МОДЕЛИРОВАНИЯ ОТТОКА КЛИЕНТОВ УСЛУГ СВЯЗИ**

*Кратко рассмотрены важнейшие функциональные модули информационной системы прогнозирования оттока клиентов телекоммуникационного предприятия. Определены основные подходы к предварительной обработке архивных данных и моделированию оттока клиентов. Заданы базовые требования для практической реализации прогностической системы. Отдельное внимание обращено на преодоление сильной коррелированности между переменными в массиве входных данных. Предложено использовать метод главных компонент, предполагающий декомпозицию входного массива на вектора счетов и нагрузок. Рассмотренный алгоритм NIPALS имеет итеративный характер. Вектор счетов, вычисленный на некоторой итерации, является соответствующей главной компонентой. Определение главных компонент дальних порядков, как правило, лишено смысла, поскольку их значения обусловлены наличием некоторой погрешности во входных данных. Указаны основные критерии для определения эффективного числа главных компонент: объясненная дисперсия и нормированное собственное значение вектора счетов. В качестве примера сформирован экспериментальный массив входных данных размера  $9 \times 2000$ , в который специально подобраны разнородные переменные (технология подключения, тип населенного пункта, скорость подключения, стоимость услуги, трафик в 1-ом месяце, трафик в 2-м месяце, трафик в 3-м месяце и др.). Отмечено, что данная методика позволяет преодолеть разнородность входной информации и сильную коррелированность переменных, а также снижает размерность входного массива. Графически показано, как число используемых главных компонент влияет на объясненную дисперсию и величину нормированного собственного значения. Все эти аспекты свидетельствуют, что данный подход перспективен для применения в прогностической системе, содержащей кластеризирующие и нейросетевые модули.*

*Прогнозирование; отток клиентов; Интернет; метод главных компонент; снижение размерности; кластеризация.*

**S.S. Alkhasov, A.N. Tselykh**

### **THE MAIN APPROACHES TO THE CREATION OF THE INFORMATION SYSTEM FOR MODELING OF TELECOMMUNICATION CLIENTS OUTFLOW**

*In the present article most important functional modules of the information system for the prediction of clients outflow from a telecom company are briefly considered. The basic approaches to pre-processing of archived data and clients outflow modeling are defined. The main requirements for the practical implementation of the prognostic system are introduced. Special attention is focused on overcoming of the strong correlation between the variables in the array of input data. It's offered to use principal components method, implying the decomposition of the input array to the score and the loading vectors. Considered algorithm NIPALS has iterative character. The score vector calculated on some iteration is the corresponding principal component. The determination of the principal components of the long-range orders doesn't have the sense typically because their values caused availability of some error in the input data. The basic criteria for definition of efficient number of principal components are*

specified: explained variance and normalized eigenvalue of score vector. As an example experimental array of the input data ( $9 \times 2000$ ) was formed. It contains selected diverse variables (technology of connection, municipality type, speed of connection, cost of service, traffic for 1st month, traffic for 2nd month, traffic for 3rd month, etc.). It's noted that these methods allow to overcome heterogeneity of input information and the strong correlation of the variables as well as to reduce the dimensionality of the input array. It's shown graphically how the number of used principal components influences on the explained variance and the normalized eigenvalue. All these aspects show that this approach is enough promising for use in the prognostic system containing modules of clustering and neurocomputing.

Prediction; clients outflow; Internet; principal component analysis; reduction of dimensionality; clustering.

**Введение.** Сегодня актуальной и коммерчески важной задачей является заблаговременное выявление клиентов услуг связи, предполагающих перейти к конкуренту, и последующее предложение им персонализированных предложений с целью их удержания. Использование исторических данных о клиентах позволяет с некоторой вероятностью предсказать последующее поведение потребителей. Соответственно требуется сформулировать подход для определения уровня лояльности клиента [1–3].

**Цель.** Получение возможности прогнозировать поведение потребителей услуг связи на основе построения математических моделей.

**Постановка задачи.** Пусть имеется массив  $\mathbf{P}$  размера  $I \times J$  архивных данных о клиентах. Его строки  $i$  ( $i = \overline{1, I}$ ) соответствуют переменным, а столбцы  $j$  ( $j = \overline{1, J}$ ) – клиентам. Параллельно существует массив  $\mathbf{T}$  целевых значений, характеризующих статус договора оказания услуги связи, где единица означает действие договора, а ноль – расторжение.

Таблица 1

Табличное представление матрицы  $\mathbf{P}$

$\mathbf{P}$	Клиент 1	...	Клиент $j$	...	Клиент $J$
Параметр 1	$p_{11}$	...	$p_{1j}$	...	$p_{1J}$
...	...	$\ddots$	...	$\ddots$	...
Параметр $i$	$p_{i1}$	...	$p_{ij}$	...	$p_{iJ}$
...	...	$\ddots$	...	$\ddots$	...
Параметр $I$	$p_{I1}$	...	$p_{Ij}$	...	$p_{IJ}$

Таблица 2

Табличное представление матрицы  $\mathbf{T}$

$\mathbf{T}$	Клиент 1	...	Клиент $j$	...	Клиент $J$
Целевое значение	$t_{11}$	...	$t_{1j}$	...	$t_{1J}$

Очевидно, что данные  $\mathbf{P}$  имеют разнородный характер, в связи с чем, должны быть подвергнуты предварительной обработке. Важнейшими методами предварительной обработки данных являются центрирование и нормирование (стандартизация) [4, 5].

Центрирование предполагает вычитание из массива данных  $\mathbf{P} = (p_{ij})$  матрицы средних значений  $\bar{\mathbf{P}} = (\bar{p}_{ij})$ .

$$p'_{ij} = p_{ij} - \bar{p}_{ij}$$

$$(\bar{p}_{ij}) = \text{colon} \left( \frac{p_{i1} + \dots + p_{ij} + \dots + p_{iJ}}{J} \right) \text{row} (u_j),$$

где  $\forall u = 1$ .

Нормирование представляет собой умножение исходного массива слева на диагональную матрицу  $\mathbf{D} = (d_{ii})$  обратных значений стандартного отклонения, рассчитанных для каждой строки (параметра) массива.

$$\begin{aligned} \mathbf{P}'' &= \mathbf{D}\mathbf{P}' \\ \mathbf{D} &= \text{diag}(\sigma_1^{-1}, \dots, \sigma_i^{-1}, \dots, \sigma_I^{-1}) \\ \sigma_i &= \sqrt{\frac{1}{J} \sum_{j=1}^J (p'_{ij} - \bar{p}_{ij})^2} \end{aligned}$$

Очевидно, что многие строки  $p_{i.}$  сильно коррелированы между собой. Простой пример: связь между скоростью соединения и входящим трафиком. Между тем применение кластеризации не является оптимальным подходом в случае сильно коррелированных переменных. Поэтому для создания новых, некоррелированных переменных [6], а также для снижения размерности данных следует применить метод главных компонент [7, 8].

В основе метода главных компонент лежит декомпозиция матрицы  $\mathbf{P}''$  на матрицу нагрузок  $\mathbf{L} = (\ell_{ik})$  и матрицу счетов  $\mathbf{S} = (s_{jk})$ . При этом не ставится задача разложить массив размера  $I \times J$  на  $I$  главных компонент, поскольку главные компоненты дальних порядков практически не несут никакой смысловой нагрузки, т.е. являются шумом  $\mathbf{E} = (e_{ij})$  [9].

$$\begin{aligned} \mathbf{P}'' &= \mathbf{L}\mathbf{S}^T + \mathbf{E} = \sum_{k=1}^K \ell_{.k} s_{.k}^T + \mathbf{E}, \\ K &\leq I, \end{aligned}$$

где  $K$  – число используемых главных компонент.

Последовательность выделения главных компонент итеративным методом NIPALS подробно описана в [10, 11]. Необходимое число главных компонент определяется посредством расчёта объясненной дисперсии.

Алгоритм метода NIPALS (рис. 1) начинается с присвоения вектор-строке  $s_0^T$  ( $J \times 1$ ) значений, соответствующих такой вектор-строке  $p''_{i.}$ , для которой сумма всех элементов максимальна [12].

$$s_0^T = \left( \arg \max_i \sum_{j=1}^J p''_{ij} \right)_{i,j=1}^{i,J}$$

Полученная вектор-строка  $s_0^T$  используется как ориентировочная на первой итерации. По смыслу она является предшественником первой главной компоненты. С помощью  $s_0^T$  формируется вектор-строка  $\ell_1$ .

$$\ell_1 = (s_0 s_0^T)^{-1} \mathbf{P}''_0 s_0^T$$

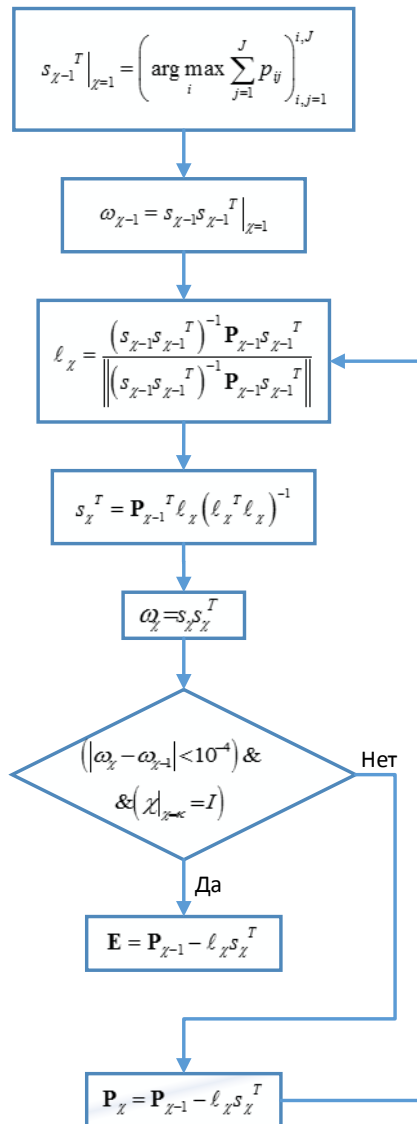


Рис. 1. Блок-схема алгоритма метода NIPALS

Далее выполняется нормирование для того, чтобы избежать вычислительных погрешностей.

$$l_1 = \frac{\ell_1}{\|\ell_1\|}$$

Затем производится уточнение приближения счетов  $S_1^T$  для данной главной компоненты путем проекции матрицы  $\mathbf{P}_0''$  на  $l_1$ .

$$s_1^T = \mathbf{P}_0''^T l_1 (l_1^T l_1)^{-1}.$$

Исходя из вычисленного  $S_1^T$  определяется собственное значение  $\omega_1$ .

$$\omega_1 = S_1 S_1^T.$$

На первой итерации  $\chi = 1$  для ориентировочной вектор-строки  $S_0^T$  также определяется собственное значение  $\omega_0$ .

После этого выполняется проверка сходимости. Модуль разности собственных значений  $\omega_\chi$  и  $\omega_{\chi-1}$  должен быть меньше  $10^{-4}$ . Покуда сходимость не достигнута, вычисляются массивы величин ошибки для каждой итерации  $\chi$  ( $\chi = \overline{1, \kappa}$ ,  $\kappa = \overline{1, I}$ , где  $\kappa$  – максимально возможное число итераций, удовлетворяющее условию  $|\omega_\chi - \omega_{\chi-1}| < 10^{-4}$ ). Эти массивы  $\mathbf{P}_1''$  имеет размерность входных данных и представляет собой разность между исходным массивом  $\mathbf{P}_0''$  и композицией векторов нагрузок  $\ell_1$  и счетов  $S_1^T$  [13].

$$\mathbf{P}_1'' = \mathbf{P}_0'' - \ell_1 S_1^T$$

Когда после  $\kappa$  итераций достигнута сходимость, рассчитывается матрица ошибок  $\mathbf{E}$ .

$$\mathbf{E} = \mathbf{P}_0'' - \ell_1 S_1^T$$

С целью выбора числа главных компонент  $K$  вычисляются нормированные собственные значения  $\lambda_k$  и объясненные дисперсии  $\eta_k$  для каждой  $k$ -ой главной компоненты [14, 15].

$$\lambda_k = \frac{\sum_{j=1}^J s_{jk}^2}{\sum_{i=1}^I \sum_{j=1}^J p_{ij}^2}$$

$$\eta_k = 1 - \frac{\sum_{i=1}^I \sum_{j=1}^J e_{ijk}^2}{\sum_{i=1}^I \sum_{j=1}^J p_{ij}^2}$$

**Экспериментальная часть.** Была рассмотрена выборка экспериментальных входных данных размера  $9 \times 2000$ , девять строк в которой соответствовали переменным (параметрам): технология подключения, тип населенного пункта, скорость подключения, стоимость услуги, трафик в 1-м месяце, трафик во 2-м месяце, трафик в 3-м месяце, разность трафика во 2-м и 1-м месяцах, разность трафика в 3-м и 2-м месяцах. Столбцы в свою очередь соответствовали 2000 клиентам. На основании расчета объясненной дисперсии и собственных значений было выбрано число главных компонент  $K = 5$ , для которого  $\eta = 0,9979$  и  $\lambda = 0,0729$  (рис. 2) [16].

Матрица счетов  $\mathbf{S}' = \mathbf{S}^T$  используется для выявления  $N$  кластеров во входных данных. В кластерном анализе известно множество методов, среди которых наибольшее распространение имеют вероятностные (к примеру, метод К-средних)

и интеллектуальные (например, сети Кохонена). В завершение данного этапа формируется вектор-строка  $\mathbf{C}$ , элемент  $c_j$  которой характеризует номер кластера  $n$  ( $n = \overline{1, N}$ ), с которым соотнесён  $j$ -й вектор-столбец массива  $\mathbf{S}'$  [17].

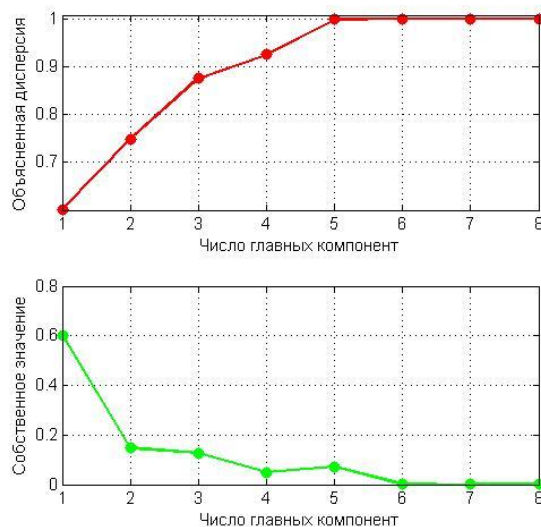


Рис. 2. Зависимости объясненной дисперсии  $\eta$  и собственных значений  $\lambda$  от числа используемых главных компонент  $K$

Вектор-строки  $\mathbf{C}$  и  $\mathbf{T}$  объединяются в матрицу размера  $2 \times J$ . Полученная матрица сортируется по номеру кластера.

$$\text{sort}_c \begin{pmatrix} \mathbf{C} \\ \mathbf{T} \end{pmatrix} = \begin{pmatrix} c_j & c_{j'} & \dots & c_{j''} \\ t_j & t_{j'} & \dots & t_{j''} \end{pmatrix},$$

где  $c_j \leq c_{j'} \leq \dots \leq c_{j''}$ .

Исходя из этого, каждому кластеру  $n$  соответствует некоторая вектор-строка  $\text{row}(\tau_{m_n})$  размера  $1 \times M_n$ , содержащая все возможные значения  $t_j$  для этого кластера.

$$\begin{pmatrix} c_j & c_{j'} & \dots & c_{j''} \\ t_j & t_{j'} & \dots & t_{j''} \end{pmatrix} = \begin{pmatrix} \overbrace{1 \quad 1 \quad 1 \quad \dots}^{n=1} & \overbrace{2 \quad \dots \quad \dots}^{n=2} \\ \underbrace{t_j \quad t_{j'} \quad t_{j''} \quad \dots}_{M_1 \text{ значений}} & \underbrace{t_{j''} \quad \dots \quad \dots}_{M_2 \text{ значений}} \end{pmatrix}$$

$\text{row}(\tau_{m_1})$                        $\text{row}(\tau_{m_2})$

Для кластера  $n$ , содержащего  $m$  ( $m = \overline{1, M}$ ) элементов при  $\forall \tau \in \{0; 1\}$ , определяется коэффициент  $\xi_n$ , характеризующий долю лояльных пользователей услуги связи в данном кластере.

$$\xi_n = \frac{1}{M_n} \sum_{m_n=1}^{M_n} \tau_{m_n}$$

Подбирается такая техника кластеризации, которая позволяет достичь значе- ний коэффициента  $\xi_n \in (0..0,2) \cup (0,8..1)$ .

Из коэффициентов  $\xi_n$  формируются вектор-строки размера  $1 \times M_n$  наподо- бие  $\text{row}(\tau_{m_n})$ , которые объединяются в вектор-строку  $\mathbf{X}$  размера  $1 \times J$ .

$$\mathbf{X} = \text{row}(x_j, x_{j'}, \dots, x_{j''})$$

$$\begin{pmatrix} c_j \\ \xi_n \end{pmatrix} = \begin{pmatrix} \overbrace{1 \quad 1 \quad 1 \quad \dots}^{g=1} & \overbrace{2 \quad \dots \quad \dots}^{g=2} \\ \underbrace{\xi_1 \quad \xi_1 \quad \xi_1 \quad \dots}_{(x_1 \dots x_{M_1})} & \underbrace{\xi_2 \quad \dots \quad \dots}_{(x_{M_1+1} \dots x_{M_1+M_2})} \end{pmatrix}$$

Затем полученный массив сортируется по номеру клиента  $j$ . Вектор-строка коэффициентов  $\xi_n$  используется в качестве целевых значений на этапе прогнози- рования.

$$\mathbf{T}' = \text{sort}_j \begin{pmatrix} c_j \\ x_j \end{pmatrix}$$

Подготовленный массив входных данных  $\mathbf{S}'$ , а также массив целевых дан- ных  $\mathbf{T}'_2$  разделяются на обучающую и тестовую подвыборки  $\tilde{\mathbf{S}}$ ,  $\tilde{\mathbf{T}}$  и  $\check{\mathbf{S}}$ ,  $\check{\mathbf{T}}$  соот- ветственно в равных долях ( $\tilde{\delta} = 0,5$  и  $\check{\delta} = 0,5$ ). Чтобы при этом сохранить идентичную однородность данных в обеих подвыборках, они могут быть разделе- ны по нечетным (odd) и четным (even) значениям  $j$ .

$$\tilde{\mathbf{S}} = \text{even}_j(s'_{kj})$$

$$\check{\mathbf{S}} = \text{odd}_j(s'_{kj})$$

$$\tilde{\mathbf{T}} = \text{even}_j(t'_{2j})$$

$$\check{\mathbf{T}} = \text{odd}_j(t'_{2j})$$

После получения преобразованных массивов входных и целевых данных осуществляется прогнозирование лояльности клиентов услуги широкополосного доступа к сети Интернет. Для решения прогностической задачи применяются не- которые виды искусственных нейронных сетей. Оценка эффективности прогноза осуществляется по величине средней квадратической ошибки  $\epsilon$ , которая не должна превосходить 0,015 [18]. Ошибки  $\tilde{\epsilon}$  и  $\check{\epsilon}$  вычисляются для обучающей и тестовой подвыборок. Причем не следует стремиться, к тому, чтобы  $\tilde{\epsilon} \rightarrow 0$ , поскольку это приведет к переобучению сети и соответственно высокому значению  $\check{\epsilon}$ . Выбор наиболее оптимальной для решения данной задачи сети достигается путём подбо- ра количества слоев и числа нейронов в них, алгоритма обучения и числа циклов обучения [19, 20].

**Заключение.** На основании рассмотрения исследований [3, 13] и анализа экспе- риментальных данных был определен набор методов для внедрения в информацио- ную систему прогнозирования оттока потребителей услуг связи. Особое внимание уделено применению метода главных компонент по отношению к специфике данной задачи, в которой входные данные изначально сильно коррелированы между собой.

Отличительной чертой данного исследования явилась как ориентированность на анализ разнородных технических и коммерческих данных о пользовательском Интернет-трафике, так и детальное выявление оптимального числа преобразованных входных переменных для последующего их анализа методами кластеризации и классификации при решении разнообразных прогностических задач.

#### БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Алхасов С.С., Целых А.Н.* Принципы построения прогностической системы для моделирования оттока клиентов услуг Интернета // Материалы Всероссийской научной конференции «Системы и модели в информационную эпоху». Ч. 1. – Таганрог: Изд-во ТТИ ЮФУ, 2014. – С. 4-6.
2. *Guazzelli A.* Predicting the future. Part 2: Predictive modeling techniques [Электронный ресурс]. – Режим доступа: <http://www.ibm.com/developerworks/library/ba-predictive-analytics2/index.html>, свободный. – Загл. с тит. экр. (дата обращения: 24.11.2014).
3. *Ефимов А.* Удержание клиента – не искусство, а прикладная наука! [Электронный ресурс]. – Режим доступа: <http://library.croc.ru/download/11342/930706a746182066f55bfe8f5800b735.pdf>, свободный. – Загл. с тит. экр. (дата обращения: 22.11.2014).
4. *Фарахутдинов Ш.Ф., Бушуев А.С.* Обработка и анализ данных социологических исследований в пакете SPSS 17.0. – Тюмень: ТюмГНГУ, 2011. – 220 с.
5. *Guazzelli A.* Predicting the future. Part 3: Create a predictive solution [Электронный ресурс]. – Режим доступа: <http://www.ibm.com/developerworks/library/ba-predictive-analytics3/index.html>, свободный. – Загл. с тит. экр. (дата обращения: 26.11.2014).
6. *Баранов С.В., Скучфина Т.П.* Новые методики и результаты исследования межрегиональной дифференциации на основе метода главных компонент // Вестник Мурманского государственного технического университета. – 2008. – Т. 11, № 2. – С. 201-210.
7. *Мокеев В.В., Томилов С.В.* Об эффективности анализа и распознавания изображений методом главных компонент и линейным дискриминантным анализом // Вестник Южно-Уральского государственного университета. Серия: Компьютерные технологии, управление, радиоэлектроника. – 2013. – Т. 13, № 3. – С. 61-70.
8. *Ким Дж.-О., Мьюллер Ч.У., Клекка У.Р. и др.* Факторный, дискриминантный и кластерный анализ: Пер. с англ. / Под ред. И.С. Енюкова. – М.: Финансы и статистика, 1989. – 215 с.
9. *Воронцов К.В.* Регрессионный анализ и метод главных компонент: Лекция [Электронный ресурс]. – Режим доступа: [http://shad.yandex.ru/lectures/machine\\_learning\\_9.xml](http://shad.yandex.ru/lectures/machine_learning_9.xml), свободный. – Загл. с тит. экр. (дата обращения: 10.12.2014).
10. *Родионова О.Е.* Интеллектуальный метод обработки многоканальных экспериментов: Дис. ... д-ра физ.-мат. наук. – М., 2008. – 272 с.
11. *Эсбенсен К.* Анализ многомерных данных. Избранные главы: Пер. с англ. / Под ред. О.Е. Родионовой. – Черноголовка: Изд-во ИПХФ РАН, 2005. – 160 с.
12. *Jolliffe I.T.* Principal Component Analysis. – 2nd edition. – NY; Berlin; Heidelberg: Springer Verlag, 2002. – 518 p.
13. *Овсянников В.П., Фефелов В.В.* Определение количества главных компонент для данных, характеризующих работу угольного предприятия, на основании анализа полной и объясненной дисперсии остатков // Современные проблемы управления производством: тезисы докладов V Международной научно-практической конференции (г. Донецк, 14–15 октября 2010 г.). – Донецк: ГВУЗ «ДонНТУ», 2010. – С. 161-163.
14. *Померанцев А.* Метод главных компонент (РСА) [Электронный ресурс]. – Режим доступа: <http://www.chemometrics.ru/materials/textbooks/pca.htm>, свободный. – Загл. с тит. экр. (дата обращения: 09.12.2014).
15. *Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д.* Прикладная статистика: Классификация и снижение размерности: Справ. изд. / Под ред. С.А. Айвазяна. – М.: Финансы и статистика, 1989. – 607 с.
16. Реализация метода главных компонент на С# [Электронный ресурс]. – Режим доступа: <http://habrahabr.ru/post/146236/>, свободный. – Загл. с тит. экр. (дата обращения: 04.12.2014).
17. *Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И.* Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP. – 2-е изд., перераб. и доп. – СПб.: БХВ-Петербург, 2007. – 384 с.
18. *Медведев В.С., Потемкин В.Г.* Нейронные сети. MATLAB 6. – М.: ДИАЛОГ-МИФИ, 2002. – 496 с.



19. *Воронцов К.В.* Нелинейная регрессия. Многослойные нейронные сети: Лекция [Электронный ресурс]. – Режим доступа: [http://shad.yandex.ru/lectures/machine\\_learning\\_10.xml](http://shad.yandex.ru/lectures/machine_learning_10.xml), свободный. – Загл. с тит. экр. (дата обращения: 24.11.2014).
20. Нейронные сети. STATISTICA Neural Networks: Методология и технологии современного анализа данных / Под ред. В.П. Боровикова. – 2-е изд., перераб. и доп. – М.: Горячая линия – Телеком, 2008. – 392 с.

## REFERENCES

1. *Alkhasov S.S., Tselykh A.N.* Printsipy postroeniya prognosticheskoy sistemy dlya modelirovaniya ottoka klientov uslug Interneta [Principles of forecasting systems for modeling churn Internet service], Materialy Vserossiyskoy nauchnoy konferentsii «Sistemy i modeli v informatsionnyu epohu» [Materials of all-Russian scientific conference "Systems and models in the information age"]. Part 1. – Taganrog: TIT SFedU, 2014, pp. 4-6.
2. *Guazzelli A.* Predicting the future, Part 2: Predictive modeling techniques [Electronic resource]. Available at: <http://www.ibm.com/developerworks/library/ba-predictive-analytics2/index.html>, free (Accessed: 24 November 2014).
3. *Efimov A.* Uderzhanie klienta – ne iskusstvo, a prikladnaya nauka! [Customer retention is not an art and applied science!] [Electronic resource]. Available at: <http://library.croc.ru/download/11342/930706a746182066f55bfe8f5800b735.pdf>, free (Accessed: 22 November 2014).
4. *Farakhutdinov Sh.F., Bushuev A.S.* Obrabotka i analiz dannykh sotsiologicheskikh issledovaniy v pakete SPSS 17.0 [Processing and analysis of data of sociological surveys in SPSS 17.0]. Tyumen: TyumGNGU, 2011, 220 p.
5. *Guazzelli A.* Predicting the future, Part 3: Create a predictive solution [Electronic resource]. Available at: <http://www.ibm.com/developerworks/library/ba-predictive-analytics3/index.html>, free (Accessed: 26 November 2014).
6. *Baranov S.V., Skufina T.P.* Novye metodiki i rezultaty issledovaniya mezhregionalnoy differentsiatsii na osnove metoda glavnkh komponent [New methods and research results of regional differentiation on the basis of principal component], *Vestnik Murmanskogo gosudarstvennogo tekhnicheskogo universiteta* [Bulletin of the Murmansk State Technical University], 2008, Vol. 11, No. 2, pp. 201-210.
7. *Mokeyev V.V., Tomilov S.V.* Ob effektivnosti analiza i raspoznavaniya izobrazheniy metodom glavnkh komponent i lineynim diskriminantnym analizom [About the effectiveness of the analysis and recognition of images by the method of principal component and linear discriminant analysis], *Vestnik Yuzhno-Uralskogo gosudarstvennogo universiteta. Seriya: Kompyuternyye tekhnologii, upravlenie, radioelektronika* [Bulletin of the South Ural State University. Computer Technologies, Automatic Control & Radioelectronics], 2013, Vol. 13, No. 3, pp. 61-70.
8. *Kim J.-O., Myuller Ch.U., Klekka U.R. et al.* Faktornyy, diskriminantnyy i klasternyy analiz [Factor, discriminant and cluster analysis]. Moscow: Finansy i statistika, 1989, 215 p.
9. *Vorontsov K.V.* Regressionnyy analiz i metod glavnkh komponentov: Lektsiya [Regression analysis and principal components: Lecture]. Available at: [http://shad.yandex.ru/lectures/machine\\_learning\\_9.xml](http://shad.yandex.ru/lectures/machine_learning_9.xml), free (Accessed 10 December 2014).
10. *Rodionova O.E.* Intellektualnyy metod obrabotki rezul'tatov mnogokanalnykh eksperimentov: Diss. dokt. fiz.-mat. nauk [Intelligent method of processing the results of multichannel experiments. Dr. phys. and math. sci. diss. Moscow, 2008, 272 p.
11. *Esbensen K.* Analiz mnogomernykh dannykh. Izbrannye glavy [Analysis of multidimensional data. Selected chapters]. Chernogolovka: IPHF RAS, 2005, 160 p.
12. *Jolliffe I.T.* Principal Component Analysis. 2<sup>nd</sup> ed. NY; Berlin; Heidelberg: Springer Verlag, 2002, 518 p.
13. *Ovsiyannikov V.P., Fefelov V.V.* Opredelenie kolichestva glavnkh komponent dlya dannykh, harakterizuyuschih rabotu ugol'nogo predpriyatiya, na osnovanii analiza polnoy i ob'yasnennoy dispersii ostatkov [The determination of the number of principal components for the data characterizing the operation of a coal mine, on the basis of detailed analysis and explained variance of the residues], *Sovremennyye problemy upravleniya proizvodstvom: tezisy dokladov V Mezhdunarodnoy nauchno-prakticheskoy konferentsii* [Modern problems of production management: abstracts of the V International scientific-practical conference] (Donetsk City, 14–15 October 2010). Donetsk: DonNTU, 2010, pp. 161-163.

14. *Pomerantsev A. Metod glavnyh komponent (PCA) [Principal component method (PCA)]. Available at: <http://www.chemometrics.ru/materials/textbooks/pca.htm>, free (Accessed 09 December 2014).*
15. *Ayvazyan S.A., Buhshtaber V.M., Enyukov I.S., Meshalkin L.D. Prikladnaya statistika: Klassifikatsiya i snizhenie razmernosti [Applied statistics: Classification and reduction of dimensionality]. Moscow: Finansyi i statistika, 1989, 607 p.*
16. *Realizatsiya metoda glavnyh komponent na C# [The implementation of the method of principal components in C#]. Available at: <http://habrahabr.ru/post/146236/>, free. (Access: 04 December 2014).*
17. *Barsegyan A.A., Kupriyanov M.S., Stepanenko V.V., Kholod I.I. Tehnologii analiza dannyih: Data Mining, Visual Mining, Text Mining, OLAP [Technology data analysis: Data Mining, Visual Mining, Text Mining, OLAP]. 2<sup>nd</sup> ed. St. Petersburg: BHV-Peterburg, 2007, 384 p.*
18. *Medvedev V.S., Potemkin V.G. Neyronnye seti. MATLAB 6 [The neural network. MATLAB 6]. Moscow: DIALOG-MIFI, 2002, 496 p.*
19. *Vorontsov K.V. Nelineynaya regressiya. Mnogosloynnye neyronnye seti: Lektsiya [Multilayer neural networks: Lecture]. Available at: [http://shad.yandex.ru/lectures/machine\\_learning\\_10.xml](http://shad.yandex.ru/lectures/machine_learning_10.xml), free (Access: 24 November 2014).*
20. *Neyronnye seti. STATISTICA Neural Networks: Metodologiya i tehnologii sovremenno go analiza dannyh [The neural network. STATISTICA Neural Networks: Methodology and technology of modern data analysis]. 2<sup>nd</sup> ed. Moscow: Goryachaya liniya – Telekom, 2008, 392 p.*

Статью рекомендовал к опубликованию д.т.н., профессор В.П. Карелин.

**Алхасов Станислав Сергеевич** – Южный федеральный университет; e-mail: [alkhasov@sfedu.ru](mailto:alkhasov@sfedu.ru); 347928, г. Таганрог, пер. Некрасовский, 44; тел.: 89081936793; кафедра информационно-аналитических систем безопасности; аспирант.

**Цельх Александр Николаевич** – e-mail: [ant@sfedu.ru](mailto:ant@sfedu.ru); кафедра информационно-аналитических систем безопасности; д.т.н., заведующий кафедрой.

**Alkhasov Stanislav Sergeevich** – Southern Federal University; e-mail: [alkhasov@sfedu.ru](mailto:alkhasov@sfedu.ru); 44, Nekrasovsky, Taganrog, 347928, Russia; phone: +79081936793; the department of information system security; postgraduate student.

**Tselykh Alexandr Nikolaevich** – e-mail: [ant@sfedu.ru](mailto:ant@sfedu.ru); the department of information system security; dr. of eng. sc., head of department.

УДК 681.5:681.3(075.8)

**Н.В. Бутырлагин, Н.Н. Прокопенко, Н.И. Чернов, В.Я. Югай**

## **СИНТЕЗ ДВОИЧНЫХ ТРИГГЕРОВ В АППАРАТЕ ЛИНЕЙНОЙ АЛГЕБРЫ\***

*Наряду с булевой алгеброй в качестве математического аппарата логического синтеза цифровых структур возможно применение и других типов алгебр. В качестве альтернативного математического аппарата предлагается линейная алгебра – алгебраическая структура  $A \rightarrow \langle A; +, \cdot, \theta; P \rangle$ , где  $P \rightarrow \langle P; +, \cdot, 0, 1 \rangle$  – поле,  $\theta$  – нулевой вектор,  $A \rightarrow \langle A; +, \cdot, \theta; P \rangle$  – векторное (линейное) пространство над полем  $P$ . Базовым множеством этой структуры является множество векторов линейного пространства соответствующей размерности. При этом множество булевых векторов является подмножеством пространства векторов линейного пространства. Достоинствами линейной алгебры являются упрощение схемотехники цифровых структур, улучшение их технических и эксплуатационных характеристик. Рассмотрены возможности применения аппарата линейной алгебры для синтеза двузначных триггерных структур. Показана линейно-алгебраическая интерпретация двузначных логических функций посредством линейных операций. В каче-*

\* Подготовлено в рамках Госзадания № 2014/38-2014 г. Минобрнауки РФ.