DOI 10.23859/1994-0637-2020-4-97-5
УДК 303.732.4

© Рапаков Г. Г., Банщиков Г. Т., Горбунов В. А., Ударатин А. В., 2020

Рапаков Георгий Германович
Кандидат технических наук, доцент,
Вологодский государственный университет
(Вологда, Россия)
E-mail: grapakov@yandex.ru

Банщиков Геннадий Трофимович
Доктор медицинских наук, врач-кардиолог,
Вологодская областная клиническая больница
(Вологда, Россия)
E-mail: vologdauzo@inbox.ru,
vol_obl_bol1@mail.ru

Горбунов Вячеслав Алексеевич
Доктор физико-математических наук, профессор,
Вологодский государственный университет
(Вологда, Россия)
E-mail: gorbunov1945@inbox.ru

Ударатин Алексей Валентинович
Кандидат технических наук, доцент,
Вологодский государственный университет
(Вологда, Россия)
ORCID 0000-0002-0372-1407
E-mail: imekn@vogu35.ru

ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ПРИ КОРРЕКЦИИ ПОВЕДЕНЧЕСКИХ ФАКТОРОВ РИСКА В ПРОФИЛАКТИКЕ СЕРДЕЧНО-СОСУДИСТЫХ ЗАБОЛЕВАНИЙ

Rapakov Georgii Germanovich
PhD in Technology, Associate Professor,
Vologda State University
(Vologda, Russia)
E-mail: grapakov@yandex.ru

Banshchikov Gennadii Trofimovich
Doctor of Medicine, M. D. Cardiologist,
Vologda region Clinical Hospital
(Vologda, Russia)
E-mail: vologdauzo@inbox.ru,
vol_obl_bol1@mail.ru

Gorbunov Vyacheslav Alekseevich
Doctor of Physics and Mathematics, Professor,
Vologda State University
(Vologda, Russia)
E-mail: gorbunov1945@inbox.ru

Udaratin Aleksey Valentinovich
PhD in Technology, Associate Professor,
Vologda State University
(Vologda, Russia)
ORCID 0000-0002-0372-1407
E-mail: imekn@vogu35.ru

USING MACHINE LEARNING METHODS FOR THE CORRECTION OF BEHAVIORAL RISK FACTORS FOR PREVENTION OF CARDIOVASCULAR DISEASES

Аннотация. Статья посвящена изучению методов машинного обучения при коррекции поведенческих факторов риска в профилактике сердечно-сосудистых заболеваний (ССЗ). Выполнен мониторинг процессов здоровьесберегающих педагогических вмешательств в структуре региональной системы организации медицинской профилактики. При помощи компьютерного моделирования построена модель правил связывания на основе метода ассоциативных правил (Association Rules). Для фактора риска

Abstract. The article describes machine learning methods in the correction of behavioral risk factors while preventing cardiovascular diseases. The monitoring of health saving educational space in the regional system of medical prevention was implemented. Applying computer modeling the authors developed a model of binding rules based on the method of association rules and suggested the set of 5 logical rules for the risk factor of arterial hypertension. Decision tree method was used to induce decision rules

(ФР) «Повышенное артериальное давление» (АД) выделен шаблон из 5 логических правил. На основе метода деревьев решений сформированы решающие правила и определена целевая аудитория влияния для адресной коррекции ФР. Выполнен анализ и получена оценка достоверности прогностической модели. Результаты анализа использованы для поддержки управленческих решений в региональной системе медицинской профилактики.

Ключевые слова: компьютерное моделирование, машинное обучение, ассоциативные правила, деревья решений, поддержка принятия решений

and identify the target group to correct risk factors and increase the quality of arterial hypertension control. The present study provided the analysis and confidence estimation of the prognostic model. The results of this analysis were used to support management decisions in the regional system of preventive medicine.

Keywords: computer modeling, machine learning, association rules, decision trees, decision support

Введение

Целенаправленное воздействие на сложный прикладной объект исследования с целью повышения эффективности его функционирования основывается на анализе системных связей с учетом закономерностей исследуемой области. Оптимизация программы развития профилактической деятельности на территориальном уровне предполагает разработку проблемно-ориентированных систем принятия решений в управлении региональным развитием с использованием современных методов анализа данных и обработки информации[1]. Актуальными проблемами являются формирование единого регионального здоровьесберегающего образовательного пространства (далее – РЗОП) на основе проектного подхода, развитие системы выявления приоритетных проблем медицинских активностей населения и организации профилактики в региональных образовательных учреждениях на основе алгоритмов и методов машинного обучения[2]. В социологических исследованиях широко применяется метод ассоциативных правил, позволяющий обнаружить связанные события, образующие поведенческий паттерн[3]. При помощи алгоритмов машинного обучения с использованием деревьев решений могут быть получены решающие правила и выделена целевая группа влияния[4]. Практическая значимость исследования определя-

[1] Рапаков Г. Г., Банщиков Г. Т. Интеллектуальный анализ данных в здравоохранении региона (на материалах Вологодской области). – Вологда: ВоГУ, 2014. – 79 с.

[2] Рапаков Г. Г., Касимов Р. А. Методы и алгоритмы машинного обучения при принятии управленческих решений в региональной системе медицинской профилактики (опыт Вологодской области). – Вологда: ВоГУ, 2014. – 143 с.

[3] Рапаков Г. Г., Касимов Р. А., Банщиков Г. Т., Горбунов В. А. Распознавание и анализ социокультурных поведенческих паттернов на основе метода ассоциативных правил // Физико-математическое моделирование систем: материалы XII Международного семинара (Воронеж, 27 июня 2014 г.): в 3 ч. – Воронеж: Воронежский государственный технический университет, 2014. – Ч. 2. – С. 155–160.

[4] Witten I. H., Frank E. Data Mining: Practical Machine Learning Tools and Techniques. – San Francisco: Morgan Kaufmann, 2005. – 525 p.; Harrington P. Machine Learning in Action. – New York: Manning Publications, 2012. – 384 p.

ется результатами компьютерного моделирования на основе методов машинного обучения при формировании РЗОП и снижения социально-экономического ущерба от преждевременной кардиоваскулярной смертности населения[1]. Цель работы состоит в изучении возможностей применения методов машинного обучения в ходе компьютерного моделирования при анализе данных медицинских активностей населения в части модификации факторов риска (далее – ФР) артериальной гипертензии (далее – АГ). Анализ данных социологического опроса позволяет определить устойчивые причинно-следственные связи в отношении населения к общепризнанным ФР, выделить шаблоны и выполнить оценку медицинских активностей, что обусловливает новизну работы. Информационно-аналитическая поддержка и адаптивная коррекция поведенческих ФР при помощи компьютерного моделирования позволяет повысить эффективность кардиоваскулярной профилактики при создании здоровьесберегающего пространства региона[2].

Основная часть

Формирование РЗОП является условием повышения эффективности реализации Государственной программы развития здравоохранения Вологодской области. Создание единой профилактической среды предполагает формирование культуры здорового образа жизни (ЗОЖ) школьников и активную вовлеченность в процесс учителей и родителей. Многоуровневая по структуре, объемам, образовательным и воспитательным ресурсам система РЗОП представляет собой масштабный по охвату населения здоровьесберегающий социально-педагогический мультипроект, в котором важное место занимает эффективное адаптивное управление развитием приоритетных направлений. Мероприятия по сохранению и укреплению здоровья граждан, реализуемые государством и обществом на национальном, региональном и муниципальном уровнях, требуют постоянно действующей системы интеллектуальной поддержки решений на всех ступенях управления, включая межведомственную. Методы машинного обучения показали свою высокую эффективность в решении задач регионального здравоохранения[3].

Артериальная гипертензия в силу своей распространенности и медико-социальной значимости находится в центре общественного внимания. В Вологодской области внедрена система раннего выявления пациентов с АГ на поликлиниче-

[1] Рапаков Г. Г., Банщиков Г. Т. Эффективность реализации областной целевой программы лечения пациентов с артериальной гипертензией на региональном уровне (опыт Вологодской области) // Экономические и социальные перемены: факты, тенденции, прогноз. – 2014. – № 5. – С. 206–221.

[2] Постановление Правительства Вологодской области от 28.10.2013 № 1112 «Об утверждении Государственной программы “Развитие здравоохранения Вологодской области” на 2014–2020 годы». – URL: http://docs.cntd.ru/document/ 422454372 (дата обращения: 16.05.2020).

[3] Рапаков Г. Г., Банщиков Г. Т. Организация системы раннего выявления больных артериальной гипертензией и доступность антигипертензивных средств в Вологодской области: опыт использования кластерного анализа // Архивъ внутренней медицины. – 2013. – № 4. – С. 16–23.

ском этапе[1]. Чем лучше выявляемость АГ на ранних стадиях, тем ниже уровень смертности от связанных с ней инфаркта миокарда и мозгового инсульта. Благодаря работе врачей первичного звена число выявленных больных АГ увеличилось. Кардиоваскулярная смертность была стабилизирована (без ее существенного снижения). Заболеваемость населения АГ (на 100 тыс. чел.) в динамике изменения для 2008–2010 гг. для муниципальных образований области представлена на рис. 1[2].

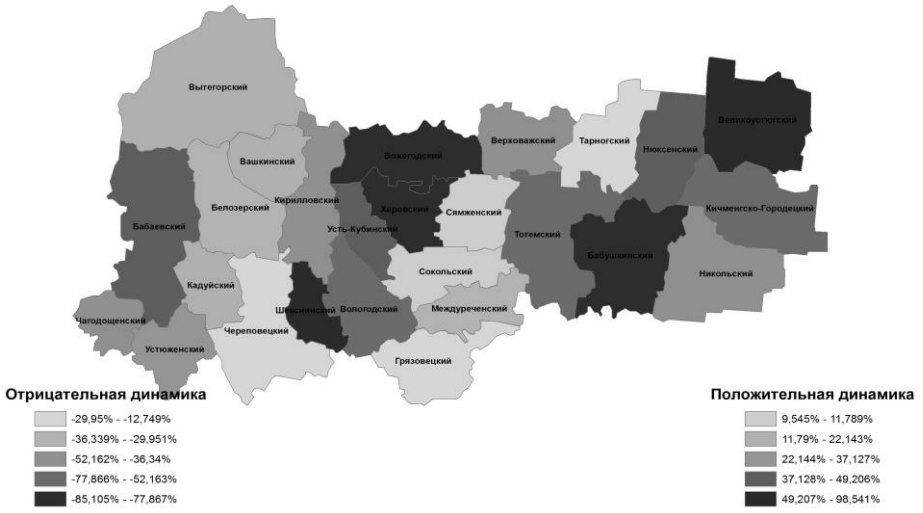


Рис. 1. Географическое распределение показателя заболеваемости артериальной гипертензией в разрезе муниципальных образований области (2008–2010 гг.)

Межведомственную работу по формированию регионального здоровьесберегающего пространства регулирует ряд нормативно-правовых актов[3]. Болезни системы

[1] Рапаков Г. Г., Банщиков Г. Т. Интеллектуальный анализ данных в здравоохранении региона (на материалах Вологодской области). – Вологда: ВоГУ, 2014. – 79 с.; Рапаков Г. Г., Банщиков Г. Т. Эффективность реализации областной целевой программы лечения пациентов с артериальной гипертензией на региональном уровне (опыт Вологодской области) // Экономические и социальные перемены: факты, тенденции, прогноз. – 2014. – № 5. – С. 206–221; Рапаков Г. Г., Банщиков Г. Т. Организация системы раннего выявления больных артериальной гипертензией и доступность антигипертензивных средств в Вологодской области: опыт использования кластерного анализа // Архивъ внутренней медицины. – 2013. – № 4. – С. 16–23.

[2] Рапаков Г. Г., Банщиков Г. Т. Визуализация показателей в задачах управления здравоохранением Вологодской области // Вузовская наука – региону: материалы XII Всероссийской научно-технической конференции (Вологда, 25 февраля 2014 г.). – Вологда: ВоГУ, 2014. – С. 61–63.

[3] Постановление Правительства Вологодской области от 28.10.2013 № 1112 «Об утверждении Государственной программы “Развитие здравоохранения Вологодской области” на 2014–2020 годы». – URL: http://docs.cntd.ru/document/ 422454372 (дата обращения: 16.05.2020); Решение Вологодской городской Думы от 29 декабря 2014 г. № 129 «Об утверждении Концепции активного долголетия на территории муниципального образования “Го-

кровообращения (далее – БСК) занимают первое место по причинам смерти в Вологодской области и за период 2009–2014 гг. в 1,18–1,28 раза превосходят показатели по РФ. В ходе реализации Государственной программы к 2020 г. ожидается снижение уровня смертности от БСК до 649,4 случаев на 100 тыс. населения. Это значение соответствует уровню 2014 г. по РФ в целом[1]. Коррекция медицинских активностей населения предполагает формирование осознанного отношения к общепризнанным факторам риска АГ и хронических неинфекционных заболеваний (далее – НИЗ), стимулирование самосохранительного поведения граждан и проведение эффективной немедикаментозной терапии[2]. Вологодский областной центр медицинской профилактики (далее – ВОЦМП) совместно с Государственным научно-исследовательским центром профилактической медицины осуществляет регулярный мониторинг по программе CINDI[3]. Для оценки влияния родителей на самосохранительное поведение детей был проведен анкетный опрос. Совокупность заполненных анкет содержит различные комбинации 134 признаков, которые появляются с разной частотой. Объем сформированной выборки $n = 274$ обеспечивает необходимую точность оценки в пределах 0,06 с доверительной вероятностью $\alpha = 0,95$.

В ходе анализа данных опроса обнаружены устойчивые причинно-следственные связи между показателями анкеты, посвященной отношению населения к ФР АГ, и получена оценка медицинских активностей родителей школьников. Для этого в исследовании были использованы модели связывания на основе метода ассоциативных правил, которые находят структуры в исходных данных и конструируют наборы правил, определяющих эти взаимосвязи[4]. В задачах машинного обучения поиск правил связывания относится к классу обучения без учителя. Необходимо выявить правила, которые представляют собой связанные группы ответов на вопросы анкеты с учетом огра-

род Вологда" на период до 2035 года "Вологда – город долгожителей"». – URL: http://docs.cntd.ru/docu-ment/438920304 (дата обращения: 16.05.2020).

[1] Постановление Правительства Вологодской области от 28.10.2013 № 1112 «Об утверждении Государственной программы "Развитие здравоохранения Вологодской области" на 2014–2020 годы». – URL: http://docs.cntd.ru/document/ 422454372 (дата обращения: 16.05.2020).

[2] Рапаков Г. Г., Банщиков Г. Т. Интеллектуальный анализ данных в здравоохранении региона (на материалах Вологодской области). – Вологда: ВоГУ, 2014. – 79 с.; Рапаков Г. Г., Касимов Р. А. Методы и алгоритмы машинного обучения при принятии управленческих решений в региональной системе медицинской профилактики (опыт Вологодской области). – Вологда: ВоГУ, 2014. – 143 с.; Рапаков Г. Г., Банщиков Г. Т. Эффективность реализации областной целевой программы лечения пациентов с артериальной гипертензией на региональном уровне (опыт Вологодской области) // Экономические и социальные перемены: факты, тенденции, прогноз. – 2014. – № 5. – С. 206–221; Рапаков Г. Г., Банщиков Г. Т. Организация системы раннего выявления больных артериальной гипертензией и доступность антигипертензивных средств в Вологодской области: опыт использования кластерного анализа // Архивъ внутренней медицины. – 2013. – № 4. – С. 16–23.

[3] Рапаков Г. Г., Касимов Р. А. Методы и алгоритмы машинного обучения при принятии управленческих решений в региональной системе медицинской профилактики (опыт Вологодской области). – Вологда: ВоГУ, 2014. – 143 с.

[4] Zhang C., Zhang S. Association Rule Mining: Models and Algorithms. – Berlin; Heidelberg: Springer-Verlag, 2002. – 248 p.

ничений. На множестве объектов X задано n бинарных признаков $F = \{f_1, ..., f_n\}$, $f_j : X \to \{0, 1\}$. Выборка $X^l = \{x_1, ..., x_l\} \subset X$ соответствует набору анкет. Рассматриваются бинарные признаки, которыми являются ответы на вопросы анкеты. Единичное значение признака $f_j(x_i) = 1$ говорит о положительном ответе на j-й вопрос в i-й анкете. При этом каждому набору признаков $\varphi \subseteq F$ ставится в соответствие предикат $\varphi(x)$, равный конъюнкции всех признаков из φ:

$$\varphi(x) = \bigwedge_{f \in \varphi} f(x), x \in X.$$

Для $\varphi(x) = 1$ признаки набора φ совместно встречаются у объекта x. Чтобы оценить связи количественно, используют показатели поддержки (support) и достоверности (confidence). Поддержку набора φ в X^l описывают функцией:

$$\nu(\varphi) = \frac{1}{l} \sum_{i=1}^{l} \varphi(x_i).$$

В целях ограничения числа правил используется параметр минимальной поддержки (minsupport) δ. Набор $\varphi \subseteq F$ называется часто встречающимся, если $\nu(\varphi) \geq \delta$. В проведенном исследовании поддержка supp – это число анкет, содержащих как условие, так и следствие относительно их общего количества. Пара непересекающихся наборов φ, $y \subseteq F$ называется ассоциативным правилом $\varphi \to y$, если выполнены следующие условия:

$$\nu(\varphi \cup y) \geq \delta; \nu(y \mid \varphi) \equiv \frac{\nu(\varphi \cup y)}{\nu(\varphi)} \geq \theta.$$

Левая часть первого из неравенств является достоверностью ассоциативного правила conf. Показатель минимальной достоверности (minconfidence) θ позволяет ограничить число правил. Достоверность используется для оценки точности правила и в исследовании представляет отношение числа анкет, содержащих как условие, так и следствие, к количеству анкет, содержащих только условие. Таким образом, $\nu(y \mid \varphi)$ рассматривается как оценка условной вероятности. Для ассоциативного правила $\varphi \to y$ справедливо: наборы φ и y совместно встречаются часто – не реже, чем в доле случаев δ; если встречается набор φ, то с частотой не менее θ встречается и набор y. Кроме того, для оценки значимости правила используется показатель Lift, представляющий собой отношение частоты появления условия среди объектов, содержащих также и следствие, к частоте появления собственно следствия. Lift рассматривают как обобщенную меру связи признаков. Метод ассоциативных правил используется для сегментации анкетируемых по поведению и анализу предпочтений[1]. Исторически для извлечения набора правил из данных широко применяется алгоритм Apriory. При этом ответы на вопросы анкеты рассматриваются как связанные причинно-следственными отношениями «из А следует С»: условие (Antecedent) → следствие (Consequent). Символ стрелки отображает правило. Ассоциатив-

[1] Асеев М. Г., Дюк В. А. Применение системы Deep Data Diver для решения задачи анализа рыночных корзин // Труды СПИИРАН. – 2004. – Вып. 2. – Т. 1. – С. 127–134.

ные правила количественно описывают связи между вопросами анкеты, соответствующие условиям и следствиям. Выделяются правила с максимальным информационным содержимым. Существуют различные рекомендации по выбору пределов поддержки и достоверности[1].

Модели деревьев решений (Decision tree) позволяют классифицировать будущие наблюдения с использованием набора решающих правил. При проведении исследования перспективным является выделение целевого сегмента лиц (носителей негативных поведенческих паттернов) при помощи деревьев решений. В здравоохранении эти модели часто применяют в ходе анализа заболеваемости[2]. Использование деревьев решений ограничено тем, что они не способны находить наиболее полные и точные правила. В работе использован алгоритм C&RT (Classification and Regression Trees). Он является одним из наиболее популярных и широко применяется в задачах классификации и регрессии автоматического анализа данных. C&RT выявляет шаблоны, отражающие системные связи и закономерности в разнородных исходных сведениях. Результатом его работы служит иерархическая структура правил в виде бинарного дерева решений, где правило является логической конструкцией вида if-then – путь от вершины до конечного узла дерева (листа). В случае бинарного дерева узел имеет двух потомков[3].

Рекурсивная дихотомия делит исходное множество на два подмножества таким образом, что записи в каждом являются более гомогенными. Очередное разбиение проводится по переменной, которая делает его лучшим. Правилом разбиения выступает индекс Gini $g(t)$. При помощи данного статистического критерия можно выполнить оценку «расстояния» между распределениями классов. Индекс $g(t)$ равен нулю для случая, когда все записи в узле относятся к одной и той же категории. Для вероятности класса j ($p(j)$) в текущем узле t индекс определяется следующим образом:

$$g(t) = 1 - \sum_j p^2(j).$$

Часто в качестве результата работы алгоритма предстает сложное дерево с большим количеством узлов и ветвей, не пригодное для интерпретации. Ценность правила уменьшается при снижении числа объектов, для которых оно справедливо. Следует избегать переобучения модели. На практике предпочтительным является результат разбиения, когда малому количеству узлов отвечает большое количество объектов. Глубину дерева ограничивают, используя оценку целесообразности дальнейшего разбиения. Однако «ранняя остановка» приводит к ухудшению классификации. В связи с этим вместо остановки используют отсечение. Точность метода определяется отношением правильно классифицированных объектов к общему чис-

[1] Zhang C., Zhang S. Association Rule Mining: Models and Algorithms. – Berlin; Heidelberg: Springer-Verlag, 2002. – 248 p.; Wu X. The Top Ten Algorithms in Data Mining. – London: Chapman & Hall, 2009. – 201 p.

[2] Рапаков Г. Г., Горбунов В. А. Интеллектуальный анализ медико-социологических данных с использованием метода Microsoft Decision Trees // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. – 2015. – № 2. – С. 130–137.

[3] Wu X. The Top Ten Algorithms in Data Mining. – London: Chapman & Hall, 2009. – 201 p.

лу объектов. При этом отсечение или замена применяются для ветвей, относительно которых ошибка распознавания не увеличивается. Из последовательности деревьев выбирается лучшее. Применяется тестирование на выборке с перекрестной проверкой. Отсечение в C&RT обеспечивает компромисс между оптимальным размером дерева и точной оценкой вероятности ошибочной классификации.

Ознакомление с литературой и ее анализ позволили выполнить сопоставление целей и методов данной работы с аналогичными источниками. Исследование М. Г. Асеева и В. А. Дюка «Применение системы Deep Data Diver для решения задачи анализа рыночных корзин»[1] посвящено применению технологии поиска ассоциативных правил, основанной на модифицированном аппарате линейной алгебры с использованием процедуры самоорганизации данных. Применяется эффект информационного структурного резонанса. Найдены высокоточные ассоциации элементов исходного множества транзакций с заданным элементом. Н. М. Жилина, А. Е. Фадеева, Г. И. Чеченин в работе «Анализ смертности населения г. Новокузнецка на основе электронной базы данных за период 1999–2007 гг.» [2] рассматривают поиск ассоциативных правил при анализе смертности. Обнаружено, что в более чем 90 % всех случаев закупорка и стеноз передней мозговой артерии распространены в пределах одного района города. В работе Г. Г. Рапакова, Р. А. Касимова, Г. Т. Банщикова, В. А. Горбунова «Распознавание и анализ социокультурных поведенческих паттернов на основе метода ассоциативных правил»[3] ассоциативные правила использованы для выделения поведенческих шаблонов в целях дальнейшей коррекции здоровьесберегающих активностей и модификации ФР БСК. Монография Г. Г. Рапакова, Р. А. Касимова «Методы и алгоритмы машинного обучения при принятии управленческих решений в региональной системе медицинской профилактики (опыт Вологодской области)» [4] обосновывает применение методов и алгоритмов интеллектуальной поддержки принятия управленческих решений при формировании РЗОП. В задачах фармако-экономического моделирования используют смешанные подходы, предполагающие совместное применение моделей Маркова и деревьев решений[5]. Для повышения качества прогноза отказов при малом количестве поломок в публикации Н. И. Шаханова, И. А. Варфоломеева, Е. В. Ершова, О. В. Юдиной «Прогнозирова-

[1] Асеев М. Г., Дюк В. А. Применение системы Deep Data Diver для решения задачи анализа рыночных корзин // Труды СПИИРАН. – 2004. – Вып. 2. – Т. 1. – С. 127–134.

[2] Жилина Н. М., Фадеева А. Е., Чеченин Г. И. Анализ смертности населения г. Новокузнецка на основе электронной базы данных за период 1999–2007 гг. // Социальные аспекты здоровья населения. – 2009. – № 3. – С. 1–11.

[3] Рапаков Г. Г., Касимов Р. А., Банщиков Г. Т., Горбунов В. А. Распознавание и анализ социокультурных поведенческих паттернов на основе метода ассоциативных правил // Физико-математическое моделирование систем: материалы XII Международного семинара (Воронеж, 27 июня 2014 г.): в 3 ч. – Воронеж: Воронежский государственный технический университет, 2014. – Ч. 2. – С. 155–160.

[4] Рапаков Г. Г., Касимов Р. А. Методы и алгоритмы машинного обучения при принятии управленческих решений в региональной системе медицинской профилактики (опыт Вологодской области). – Вологда: ВоГУ, 2014. – 143 с.

[5] Крысанов И. С. Введение в фармакоэкономическое моделирование // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. – 2008. – Т. 1. – № 1. – С. 7–9.

ние отказов оборудования в условиях малого количества поломок»[1] рассмотрен метод деревьев решений (Random Forest). На основе исходной обучающей выборки с возвращением алгоритм строит большое количество решающих деревьев. Для настройки использованы данные об отказах, в ходе обучения модели – сведения, которые соответствуют нормальному режиму работы оборудования. Отказы и аномалии выявлены при помощи разности показаний фактического и прогнозного значений нормального сигнала в последующий интервал времени.

Особенностью предметной области исследования является широкая распространенность поведенческих ФР НИЗ. Модель связывания позволяет эффективно и быстро определить устойчивые причинно-следственные отношения между социометрическими показателями многомерного пространства. Оценочные показатели порогов поддержки и достоверности были предложены при постановке задачи в сотрудничестве со специалистами ВОЦМП, исходя из требований анализа системных связей медицинских активностей населения в части модификации ФР АГ. При этом в задаче оценки медицинских активностей населения в качестве следствия рассматривался отрицательный ответ на вопрос «Является ли повышенное кровяное давление значимым фактором риска?» Утверждение «ФР – повышенное АД = нет» ложно, оно свидетельствует о том, что для интервьюируемого данный ФР не важен. Из всего множества условий, отобранных моделью связывания CARMA при варьировании параметров support (от 6 % до 100 %) и confidence (от 60 % до 100 %), были выделены пять, имеющих отношение к следствию «ФР – повышенное АД = нет», с поддержкой более 90 % и достоверностью выше 97 %, представляющих интерес для анализа (см. таблицу).

Таблица

Ассоциативная модель связанных событий

Consequent (следствие): ФР – повышенное АД = нет		
Antecedent (условие)	Поддержка (supp.), %	Достоверность (conf.), %
ФР – избыточный вес = нет	98,54	97,78
Благополучие – никто = нет	95,99	97,72
Говорите на другие темы о здоровье = не обсуждаю	94,53	97,68
Программа ЗОЖ в школе нужна = да	90,88	97,59
Благополучие – собственные действия = нет	90,15	98,38

С точки зрения организации профилактической работы из числа распознанных как положительные правил можно рассматривать ответы на вопросы: «Нужна ли программа здорового образа жизни в школе?» – ответ «Да» и «Кто или что может повлиять на благополучие и развитие населения?» – вариант «Никто не может повлиять» – ответ «Нет».

[1] Шаханов Н. И., Варфоломеев И. А., Ершов Е. В., Юдина О. В. Прогнозирование отказов оборудования в условиях малого количества поломок // Вестник Череповецкого государственного университета. – 2016. – № 6 (75). – С. 36–41.

Однако остальные три правила являются ложными стереотипами узнавания, свидетельствующими о том, что в сознании анкетируемого искажено восприятие не только следствия в рамках шаблона, но и его условий. Так, на вопрос «Говорите на другие темы о здоровье?» с поддержкой 94,5 % ответом является «Не обсуждаю». Между тем в анкете были перечислены лишь базовые темы ЗОЖ и опущены вопросы формирования духовных качеств, нравственного здоровья, самореализации и личностного роста. Поддержка утверждения, что избыточный вес не является значимым ФР, составляет 98,5 %. Однако суммарный вклад двух из семи ведущих факторов риска сердечно-сосудистых заболеваний (АГ и избыточной массы тела) обусловливает около 48 % преждевременной смертности[1].

Для вопроса «Кто или что может повлиять на благополучие и развитие населения?» вариант «Мои собственные действия» – ответ «Нет» имеет поддержку в 90,1 % случаев. Это свидетельствует о глубокой пассивности подавляющей части населения. Отрицательные социальные манипуляции по отношению к индивидууму при отсутствии нравственных идеалов в условиях общественной нестабильности ведут к безразличию к себе и окружающим – форме апатии отчаяния в условиях управляемого хаоса.

Выделенные ассоциации образуют набор – поведенческий паттерн. Этот шаблон, наряду с многочисленными другими, выполняет функции регулирования в жизни анкетируемых. Наборы транслируются респондентами в свое окружение. Системный подход предусматривает работу с компонентами шаблона как с единым целым, повышая рентабельность мероприятий по профилактике заболеваний. Визуализация связей расширенного набора ассоциативных правил для переменных с поддержкой выше 90 % отражена на рис. 2 при помощи сетевого графа.

Иерархическая структура решающих правил в виде дерева решений представлена на рис. 3. Целевым параметром выступает принадлежность к группе респондентов, выявленных моделью связывания, построенной с использованием ассоциативных правил (см. таблицу).

[1] Рапаков Г. Г., Банщиков Г. Т. Интеллектуальный анализ данных в здравоохранении региона (на материалах Вологодской области). – Вологда: ВоГУ, 2014. – 79 с.; Рапаков Г. Г., Касимов Р. А. Методы и алгоритмы машинного обучения при принятии управленческих решений в региональной системе медицинской профилактики (опыт Вологодской области). – Вологда: ВоГУ, 2014. – 143 с.; Рапаков Г. Г., Банщиков Г. Т. Эффективность реализации областной целевой программы лечения пациентов с артериальной гипертензией на региональном уровне (опыт Вологодской области) // Экономические и социальные перемены: факты, тенденции, прогноз. – 2014. – № 5. – С. 206–221; Рапаков Г. Г., Банщиков Г. Т. Организация системы раннего выявления больных артериальной гипертензией и доступность антигипертензивных средств в Вологодской области: опыт использования кластерного анализа // Архивъ внутренней медицины. – 2013. – № 4. – С. 16–23.

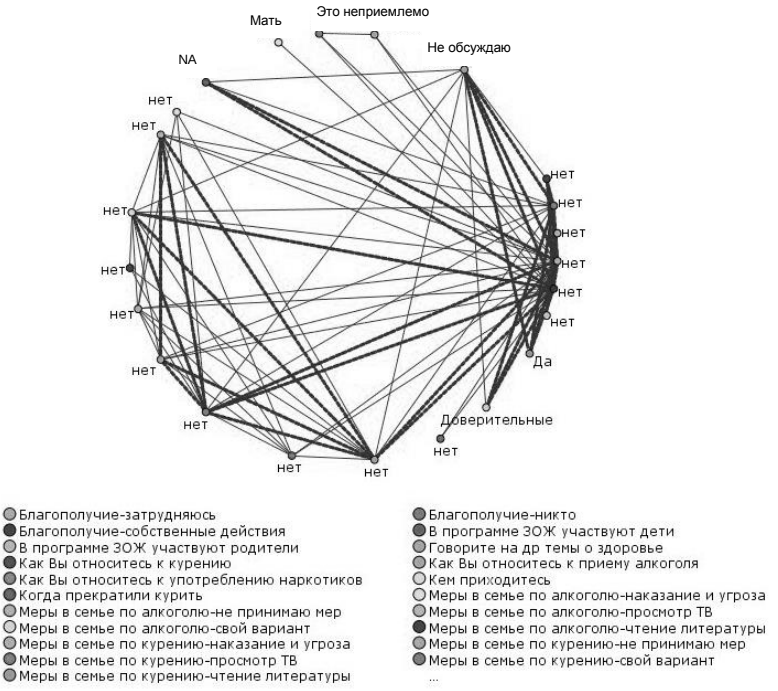


Рис. 2. Визуализация связей расширенного набора ассоциативных правил при помощи сетевого графа

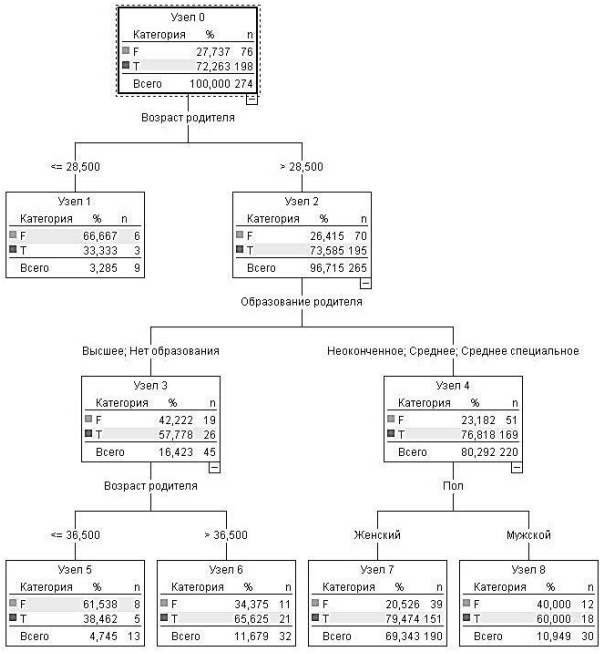


Рис. 3. Иерархическая структура решающих правил в виде дерева решений

Аудитория влияния при этом представлена лицами в возрасте более 28,5 лет с незаконченным, средним и средним специальным образованием, женского пола. Объем выделенной целевой аудитории составляет 69 %. Для сужения целевой группы необходимо включение дополнительной персональной информации в опросный лист.

Выполнена оценка влияния переменных набора данных на результаты классификации. Определяющий вклад в формирование ложных поведенческих стереотипов, увеличивающих воздействие факторов риска АГ, существенно снижающих эффективность профилактики и качество реализации РЗОП, а также увеличивающих социально-экономическое бремя БСК, вносит образование родителя (51 %). Для целевой группы – это незаконченное, среднее и среднее специальное образование. Оценка значимости возраста родителя составляет 24 %. Пол и степень родства определяют 16 % и 9 % влияния соответственно.

Поскольку суммарный вклад двух факторов риска (АГ и избыточной массы тела) обусловливает около 48 % преждевременной смертности, можно утверждать, что низкий образовательный статус не только снижает качество жизни и блокирует возможность самоактуализации личности, но и вносит существенный вклад в преждевременную кардиоваскулярную смертность.

Верификация модели позволяет выполнить проверку достоверности полученного знания. Точность классификации прогностической модели была определена при помощи процедуры анализа на основе таблицы сопряженности дерева решений. Результаты показывают, что 75,55 % значений, предсказанных моделью, соответствуют фактическим значениям, что вполне достаточно для практических приложений.

Выводы

Для решения задачи обнаружения связанных событий в данных социологического опроса был использован метод машинного обучения CARMA Algorithms. На основе ассоциативных правил выявлены психосоциальные характеристики, сопоставленные с поведенческими факторами риска НИЗ. Для фактора риска «Повышенное артериальное давление» выделен шаблон из пяти логических правил с поддержкой не менее 90 % и достоверностью не ниже 97 %, который отражает закономерности исследуемой области профилактической медицины. Для подавляющей части респондентов характерна социальная пассивность. Визуализация связей расширенного набора ассоциативных правил выполнена на основе сетевого графа. При помощи метода деревьев решений C&RT Algorithms сформированы решающие правила и выявлен целевой сегмент лиц – носителей негативных социокультурных поведенческих установок – для последующей коррекции их здоровьесберегающих активностей и модификации ФР. Целевая группа представлена респондентами в возрасте более 28,5 лет с незаконченным, средним и средним специальным образованием, женского пола. Объем выделенной целевой аудитории составляет 69 %. Определяющий вклад (более 50 %) в формирование выявленного негативного социокультурного поведенческого паттерна вносит образование родителя. Его низкий уровень существенно влияет на преждевременную кардиоваскулярную смертность.

Дальнейшие перспективы работы связаны с геоинформационным моделированием заболеваемости и смертности от БСК на основе метода пространственных ассо-

циативных правил. Результаты исследования использованы для интеллектуальной поддержки принятия решений в задачах повышения эффективности профилактики, коррекции здоровьесберегающих активностей и факторов риска в ходе выполнения региональной программы развития здравоохранения.

Литература

Асеев М. Г., Дюк В. А. Применение системы Deep Data Diver для решения задачи анализа рыночных корзин // Труды СПИИРАН. – 2004. – Вып. 2. – Т. 1. – С. 127–134.

Жилина Н. М., Фадеева А. Е., Чеченин Г. И. Анализ смертности населения г. Новокузнецка на основе электронной базы данных за период 1999–2007 гг. // Социальные аспекты здоровья населения. – 2009. – № 3. – С. 1–11.

Крысанов И. С. Введение в фармакоэкономическое моделирование // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. – 2008. – Т. 1. – № 1. – С. 7–9.

Постановление Правительства Вологодской области от 28.10.2013 № 1112 «Об утверждении Государственной программы “Развитие здравоохранения Вологодской области” на 2014–2020 годы». – URL: http://docs.cntd.ru/document/422454372 (дата обращения: 16.05.2020).

Рапаков Г. Г., Банщиков Г. Т. Визуализация показателей в задачах управления здравоохранением Вологодской области // Вузовская наука – региону: материалы XII Всероссийской научно-технической конференции (Вологда, 25 февраля 2014 г.). – Вологда: ВоГУ, 2014. – С. 61–63.

Рапаков Г. Г., Банщиков Г. Т. Интеллектуальный анализ данных в здравоохранении региона (на материалах Вологодской области). – Вологда: ВоГУ, 2014. – 79 с.

Рапаков Г. Г., Банщиков Г. Т. Организация системы раннего выявления больных артериальной гипертензией и доступность антигипертензивных средств в Вологодской области: опыт использования кластерного анализа // Архивъ внутренней медицины. – 2013. – № 4. – С. 16–23.

Рапаков Г. Г., Банщиков Г. Т. Эффективность реализации областной целевой программы лечения пациентов с артериальной гипертензией на региональном уровне (опыт Вологодской области) // Экономические и социальные перемены: факты, тенденции, прогноз. – 2014. – № 5. – С. 206–221.

Рапаков Г. Г., Горбунов В. А. Интеллектуальный анализ медико-социологических данных с использованием метода Microsoft Decision Trees // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. – 2015. – № 2. – С. 130–137.

Рапаков Г. Г., Касимов Р. А. Методы и алгоритмы машинного обучения при принятии управленческих решений в региональной системе медицинской профилактики (опыт Вологодской области). – Вологда: ВоГУ, 2014. – 143 с.

Рапаков Г. Г., Касимов Р. А., Банщиков Г. Т., Горбунов В. А. Распознавание и анализ социокультурных поведенческих паттернов на основе метода ассоциативных правил // Физико-математическое моделирование систем: материалы XII Международного семинара (Воронеж, 27 июня 2014 г.): в 3 ч. – Воронеж: Воронежский государственный технический университет, 2014. – Ч. 2. – С. 155–160.

Решение Вологодской городской Думы от 29 декабря 2014 г. № 129 «Об утверждении Концепции активного долголетия на территории муниципального образования “Город Вологда” на период до 2035 года “Вологда – город долгожителей”». – URL: http://docs.cntd.ru/document/438920304 (дата обращения: 16.05.2020).

Шаханов Н. И., Варфоломеев И. А., Ершов Е. В., Юдина О. В. Прогнозирование отказов оборудования в условиях малого количества поломок // Вестник Череповецкого государственного университета. – 2016. – № 6 (75). – С. 36–41.

Harrington P. Machine Learning in Action. – New York: Manning Publications, 2012. – 384 p.

Witten I. H., Frank E. Data Mining: Practical Machine Learning Tools and Techniques. – San Francisco: Morgan Kaufmann, 2005. – 525 p.

Wu X. The Top Ten Algorithms in Data Mining. – London: Chapman & Hall, 2009. – 201 p.

Zhang C., Zhang S. Association Rule Mining: Models and Algorithms. – Berlin; Heidelberg: Springer-Verlag, 2002. – 248 p.

References

Aseev M. G., Diuk V. A. Primenenie sistemy Deep Data Diver dlia resheniia zadachi analiza rynochnykh korzin [Deep Data Diver application in market baskets analysis]. *Trudy SPIIRAN* [SPIIRAS Proceedings], 2004, no. 2, vol. 1, pp. 127–134.

Zhilina N. M., Fadeeva A. E., Chechenin G. I. Analiz smertnosti naseleniia g. Novokuznetska na osnove elektronnoi bazy dannykh za period 1999–2007 gg. [The death-rates analysis of Novokuznetsk population on the basis of an electronic database for the period 1999–2007]. *Sotsial'nye aspekty zdorov'ia naseleniia* [Social aspects of population health], 2009, no. 3, pp. 1–11.

Krysanov I. S. Vvedenie v farmakoekonomicheskoe modelirovanie [Introduction into pharmacoeconomic modelling]. *Farmakoekonomika. Sovremennaia farmakoekonomika i farmakoepidemiologiia* [Farmakoekonomika. Modern Pharmacoeconomic and Pharmacoepidemiology], 2008, vol. 1, no. 1, pp. 7–9.

Postanovlenie Pravitel'stva Vologodskoi oblasti ot 28.10.2013 № 1112 "Ob utverzhdenii Gosudarstvennoi programmy 'Razvitie zdravookhraneniia Vologodskoi oblasti' na 2014–2020 gody" [Decree of the Government of Vologda region of 28.10.2013 № 1112 "On the approval of the state programme 'The development of healthcare in Vologda region for the period 2014–2020'"]. Available at: http://docs.cntd.ru/document/422454372 (accessed: 16.05.2020).

Rapakov G. G., Banshchikov G. T. Vizualizatsiia pokazatelei v zadachakh upravleniia zdravookhraneniem Vologodskoi oblasti [Visualization of indicators in health management tasks in Vologda region]. *Vuzovskaia nauka – regionu: materialy XII Vserossiiskoi nauchno-tekhnicheskoi konferentsii* [University science – region: proceedings of the 12th all-Russian scientific and technical conference (Vologda, 25 February, 2014)].Vologda: Vologodskii gosudarstvennyi universitet, 2014, pp. 61–63.

Rapakov G. G., Banshchikov G. T. *Intellektual'nyi analiz dannykh v zdravookhranenii regiona (na materialakh Vologodskoi oblasti)* [Data mining in region public health (by the material of Vologda region)]. Vologda: Vologodskii gosudarstvennyi universitet, 2014. 79 p.

Rapakov G. G., Banshchikov G. T. Organizatsiia sistemy rannego vyiavleniia bol'nykh arterial'noi gipertenziei i dostupnost' antigipertenzivnykh sredstv v Vologodskoi oblasti: opyt ispol'zovaniia klasternogo analiza [Organization of early detection system for patients with arterial hypertension and availability of antihypertensive drugs in Vologda region: experience of cluster analysis]. *Arkhiv" vnutrennei meditsiny* [The Russian Archives of Internal Medicine], 2013, no. 4, pp. 16–23.

Rapakov G. G., Banshchikov G. T. Effektivnost' realizatsii oblastnoi tselevoi programmy lecheniia patsientov s arterial'noi gipertenziei na regional'nom urovne (opyt Vologodskoi oblasti) [Efficiency of implementation of the regional target program for the treatment of patients with arterial hypertension at the regional level (experience of the Vologda Oblast)]. *Ekonomicheskie i sotsial'nye peremeny: fakty, tendentsii, prognoz* [Economic and Social Changes: Facts, Trends, Forecast], 2014, no. 5, pp. 206–221.

Rapakov G. G., Gorbunov V. A. Intellektual'nyi analiz mediko-sotsiologicheskikh dannykh s ispol'zovaniem metoda Microsoft Decision Trees [Intelligent analysis of medical-sociological data

using Microsoft Decision Trees algorithm]. *Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriia: Sistemnyi analiz i informatsionnye tekhnologii* [Proceedings of Voronezh State University. Series: Systems analysis and information technologies], 2015, no. 2, pp. 130–137.

Rapakov G. G., Kasimov R. A. *Metody i algoritmy mashinnogo obucheniia pri priniatii upravlencheskikh reshenii v regional'noi sisteme meditsinskoi profilaktiki (opyt Vologodskoi oblasti)* [Methods and algorithms of machine learning for decision-making in the regional system of medical prevention (experience of Vologda region)]. Vologda: Vologodskii gosudarstvennyi universitet, 2014. 143 p.

Rapakov G. G., Kasimov R. A., Banshchikov G. T., Gorbunov V. A. Raspoznavanie i analiz sotsiokul'turnykh povedencheskikh patternov na osnove metoda assotsiativnykh pravil [Recognition and analysis of behavioral sociocultural patterns by association rules technique]. *Fiziko-matematicheskoe modelirovanie sistem: materialy XII mezhdunarodnogo seminara (Voronezh, 27 iiunia 2014 g.: v 3 ch.)* [Physico-mathematical system modeling: XII international seminar materials (Voronezh, 27 June 2014): in 3 parts]. Voronezh: Voronezhskii gosudarstvennyi tekhnicheskii universitet, 2014, part 2, pp. 155–160.

Reshenie Vologodskoi gorodskoi Dumy ot 29 dekabria 2014 g. № 129 "Ob utverzhdenii Kontseptsii aktivnogo dolgoletiia na territorii munitsipal'nogo obrazovaniia 'Gorod Vologda' na period do 2035 goda 'Vologda – gorod dolgozhitelei'" [Vologda – a city of long-livers: the concept for active longevity on the territory of the municipal formation "City of Vologda" for the period till 2035: approved by the decision of the Vologda City Duma of December 29, 2014. No. 129]. Available at: http://docs.cntd.ru/document/438920304 (accessed: 16.05.2020).

Shakhanov N. I., Varfolomeev I. A., Ershov E. V., Iudina O. V. Prognozirovanie otkazov oborudovaniia v usloviiakh malogo kolichestva polomok [Forecasting equipment failures under the conditions of a small number of breakdowns]. *Vestnik Cherepovetskovo gosudarstvennovo universiteta* [Bulletin of the Cherepovets State University], 2016, no. 6 (75), pp. 36–41.

Harrington P. *Machine Learning in Action*. New York: Manning Publications, 2012. 384 p.

Witten I. H., Frank E. *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco: Morgan Kaufmann, 2005. 525 p.

Wu X. *The Top Ten Algorithms in Data Mining*. London: Chapman & Hall, 2009. 201 p.

Zhang C., Zhang S. *Association Rule Mining: Models and Algorithms*. Berlin; Heidelberg: Springer-Verlag, 2002. 248 p.

Для цитирования: *Рапаков Г. Г., Банщиков Г. Т., Горбунов В. А., Ударатин А. В.* Использование методов машинного обучения при коррекции поведенческих факторов риска в профилактике сердечно-сосудистых заболеваний // Вестник Череповецкого государственного университета. – 2020. – № 4 (97). – С. 54–68. DOI: 10.23859/1994-0637-2020-4-97-5

For citation: Rapakov G. G., Banshchikov G. T., Gorbunov V. A., Udaratin A. V. Using machine learning methods for the correction of behavioral risk factors for prevention of cardiovascular diseases. *Bulletin of the Cherepovets State University*, 2020, no. 4 (97), pp. 54–68. DOI: 10.23859/1994-0637-2020-4-97-5