

Когнитивная визуализация классифицирующих правил, извлеченных из данных, на основе модели бинарной решающей матрицы

В. И. Пименов^а, доктор техн. наук, профессор, orcid.org/0000-0002-7228-3009, v_pim@mail.ru

М. В. Воронов^б, доктор техн. наук, профессор, orcid.org/0000-0002-0691-7245

И. В. Пименов^в, канд. техн. наук, доцент, orcid.org/0000-0002-1954-6463

^аСанкт-Петербургский государственный университет промышленных технологий и дизайна, Б. Морская ул., 18, Санкт-Петербург, 191186, РФ

^бМосковский государственный психолого-педагогический университет, Открытое ш., 24, стр. 27, Москва, 107143, РФ

^вГосударственный университет морского и речного флота им. адмирала С. О. Макарова, Двинская ул., 5/7, Санкт-Петербург, 198035, РФ

Введение: огромные объемы данных генерируются в интернет-пространстве и из внутренней информации организаций. Для получения набора выходных данных с четкой структурой, их разделения на значимые части и построения классифицирующих правил применяются методы машинного обучения. Большая часть индуктивных методов моделирует в многомерном пространстве промежуточные и высокоуровневые абстрактные категории, которые сложно интерпретировать.

Цель: разработка модели машинного обучения в виде «белого ящика», объясняющего выбранное решение при помощи обычных продукционных правил, и когнитивных визуализаторов для характеристики классов объектов. **Методы:** построение бинарной решающей матрицы, предоставляющей информацию о комплексе значений выбранных свойств объектов, которые имплицитно заданные классы. **Результаты:** бинарная решающая матрица формируется автоматически по результатам кластерного и дискриминантного анализов. Процесс обучения ограничивается выставлением пороговых уровней и единичных элементов матрицы, что обеспечивает возможность семантической интерпретации решающего правила. Процедура распознавания сводится к поэлементной конъюнкции тех строк матрицы, которые определяются атрибутами объекта. Выделенный единичный элемент указывает номер класса. Для интерпретации правила разработан универсальный алгоритм обработки бинарной матрицы, использующий вводимые пользователем значения признаков. Размерность просмотренного пространства определяется кольцами мерности на визуализаторе распознавания. Азимут иницированной ячейки диаграммы, имеющей наибольшую размерность, указывает принадлежность объекта с установленными признаками целевому классу. Для характеристики классов разработаны визуализаторы, демонстрирующие как их отличительные свойства, так и свойства, общие для нескольких классов. Во многих случаях распознавание типа объекта останавливается при существенно меньшей глубине просмотренного пространства признаков, чем при полном переборе. **Практическая значимость:** предложенные методы когнитивного анализа и визуализации данных выполняют не только классификацию данных, упорядочивание и отбор значимых признаков, но и построение правил, раскрывающих причинно-следственные связи между комбинацией факторов и типом принимаемого решения.

Ключевые слова — большие данные, когнитивные технологии, многомерный анализ данных, машинное обучение, решающее правило, бинарная решающая матрица, визуализатор.

Для цитирования: Пименов В. И., Воронов М. В., Пименов И. В. Когнитивная визуализация классифицирующих правил, извлеченных из данных, на основе модели бинарной решающей матрицы. *Информационно-управляющие системы*, 2019, № 6, с. 2–11. doi:10.31799/1684-8853-2019-6-2-11

For citation: Pimenov V. I., Voronov M. V., Pimenov I. V. The cognitive visualization of classifying rules extracted from data based on binary solver matrix model. *Informatsionno-upravlyaiushchie sistemy* [Information and Control Systems], 2019, no. 6, pp. 2–11 (In Russian). doi:10.31799/1684-8853-2019-6-2-11

Введение

Умение эффективно обрабатывать неструктурированные данные — один из факторов успеха в науке, медицине, бизнесе и других сферах, где генерируются большие объемы информации. Целью изучения и анализа больших данных является получение организованного набора выходных данных с четкой структурой и обнаружение закономерностей.

Системы поддержки принятия решений, использующие большие объемы гетерогенных много-

мерных данных, находят применение при планировании стратегического развития предприятия, оценке текущей ситуации, стратегических альтернатив и интегральной эффективности управления корпорацией. Когнитивный анализ и визуализация данных используются для выявления техносферных опасностей, сетевого оперативного мониторинга, категоризации сочетания симптомов как некоторого заболевания, стратификации вузов по стоимости обучения и качеству обучения, представления кредитных историй клиентов банка, контроля состояния сложного технического объекта [1–4].

Когнитивный анализ данных включает изучение и внедрение методов, с помощью которых эксперт оценивает текущую ситуацию, обнаруживает влияние скрытых факторов, описывает выявленные закономерности, использует их для получения новых знаний, принятия решений и прогнозирования развития ситуации. Использование когнитивных технологий позволяет ослабить усилия по развитию методов хранения информации, заменив их методами автоматизированного построения баз знаний [5–6].

Основной проблемой когнитивного анализа является кластеризация данных, их разделение на значимые части, определение законов, правил и алгоритмов, по которым происходит отнесение объекта к определенному классу. Реализованные в разрабатываемых методах принципы должны быть по возможности когнитивно валидными, соответствовать потенциальной когнитивной структуре [7].

Для успешного решения различных задач анализа данных, от агрегации, группировки объектов до характеристики классов объектов с помощью их атрибутов, модель должна быть универсальной и основанной на способах обработки информации человеком. Универсальный психофизиологический механизм познания [8] использует способность человека оценивать сходство между объектами, его стремление к компактному представлению содержательно близких объектов и обобщению их описаний в терминах выбранной меры подобия.

Вместе с тем во многих ситуациях эксперт в состоянии выполнить прогноз поведения сложной системы, если на нее воздействует не более 3–5 слабо зависимых факторов внешней среды. Если входные переменные взаимосвязаны, то их влияние не способен учесть даже опытный специалист [9].

Задачи количественного анализа предшествующего опыта, отраженного в рядах накопленных данных, формирование множества вариантов решений и прогностических сценариев развития событий должны решаться с применением систем когнитивного анализа данных.

Основные требования к таким системам:

- обработка данных сверхбольшого объема;
- разнородность данных;
- значительная глубина анализа;
- интерпретируемость результатов анализа;
- доступность и простота инструментария.

Инструментальной основой для извлечения закономерностей из информации об окружающем мире являются методы машинного обучения и многомерного анализа данных, применимые для широкого спектра объектов исследования. Принадлежность объекта к одному из альтернативных классов определяется с помощью построенного в процессе обучения алгоритма распознавания посредством геометрического представления.

Проблемой остается построение модели «белого ящика», обеспечивающей семантическую интерпретацию решений на основе формальных правил.

Обучение «белого ящика»

Трудность вербализации результатов глубокого обучения нейросетевой модели и объяснений, почему она приняла то или иное решение, несмотря на широкий спектр задач, решаемых нейронными сетями, связана с подходом «черного ящика» [10–12].

Интерпретируемые результаты показывают деревья решений. Однако автоматически построенные для больших данных деревья представляют собой громоздкие конструкции, в которых большинство промежуточных концептов являются абстрактными группами, не поддающимися семантической интерпретации. Построение же оптимального дерева решений, необходимого для получения обобщающих правил, является NP-полной задачей [13].

Опираясь на гипотезу, согласно которой закономерности обработки информации человеком выражаются нейронными кодами, а сущности предметной области могут быть представлены кодовыми обозначениями, примем построение дискретных моделей знаний как основу когнитивной обработки многомерных данных. Связь между кодовыми комбинациями значений признаков и возможными классами в явной форме устанавливают таблицы принятия решений и таблицы истинности [14].

В первом случае таблица составляется для всех комбинаций многоградационных признаков. Обученные алгоритмически на основе больших данных таблицы являются избыточными и требуют дополнительной редукции набора признаков [15]. Неоднозначность, при которой число найденных классов может быть больше одного, объясняется интерпретацией классов как решений о последовательности действий. Во втором случае таблицы ограничены применением булевых переменных и двумя классами объектов.

Целесообразным является использование модели «белого ящика», представляющей результаты машинного обучения в виде сокращенной бинарной решающей матрицы (БРМ) [16] и диаграммы многовариантного решения, объясняющих выбранный вариант при помощи булевой логики. Процедуры сокращения позволяют исключить построение промежуточных концептов для реализации последовательных методов поиска [17], а процесс распознавания выполнить в момент получения значений для достаточного числа признаков объекта.

Исходный массив данных исследуется в глобальном пространстве измеримых свойств нако-

пленных объектов. Данные о многомерных объектах $\omega_i, i = \overline{1, n}$, представляются вектором значений $(x_{i1}, \dots, x_{ij}, \dots, x_{iN_a})$ признаков априорного словаря $\mathbf{X}_a, \mathbf{X}_a = \{X_j | j = \overline{1, N_a}\}$, содержащего перечень измеримых свойств объектов, и сведениями о принадлежности объектов обучающей выборки к одному из установленных попарно непересекающихся классов $\Omega_m, \Omega_m \subset \Omega, \Omega = \Omega_1 \cup \Omega_2 \cup \dots \cup \Omega_M$, которые описываются классификационной переменной Y .

Бинарная решающая матрица формируется автоматически по данным, полученным в ходе кластеризации данных и дискриминантного анализа (рис. 1) [16]. БРМ выполняет адаптивное квантование признакового пространства и предоставляет информацию о кластерах в виде интерпретируемых бинарных кодов. В ней содержатся сведения о комбинации тех значений выбранных информативных признаков предъявляемых объектов, которые имплицитно выделяют классы.

Процесс обучения включает в себя квантование признаков методом, обеспечивающим разделение всех непересекающихся по каждому при-

знаку классов с использованием минимального числа порогов, и их отбор методами целочисленного программирования, при котором сокращенное описание классов сохраняет дискриминирующую силу исходного признакового пространства. Координаты элементов матрицы являются декартовым произведением множества градаций $\{t_j | j = \overline{1, N_p}\}$ сокращенного (рабочего) словаря признаков $\mathbf{X}_p = \{X_j | j = \overline{1, N_p}\}$ на множество классов $\{\Omega_1, \Omega_2, \dots, \Omega_M\}$.

В отличие от таблиц принятия решений, сокращенная БРМ предоставляет в явном виде для каждой градации признака бинарную кодовую комбинацию классов, что позволяет алгоритмизировать интерпретацию принимаемого решения и визуализировать найденный образ.

Параметрическая идентификация матрицы заключается в назначении единичных значений элементам матрицы x_{ij} , если у объектов m -го класса установлен бинарный признак либо значения количественного признака X_j принадлежат интервалу с номером t :

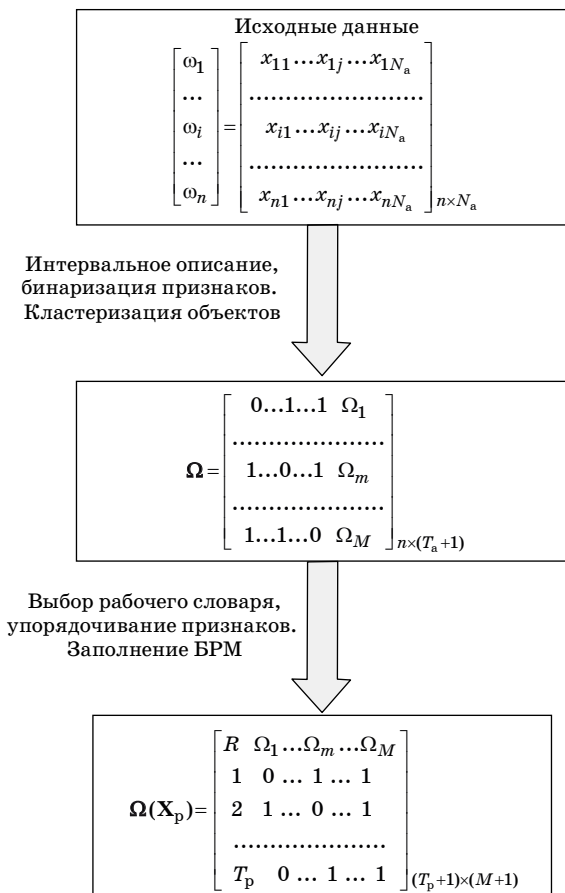
$$x_{ij}(m) = \begin{cases} 1, \exists \omega \in \Omega_m, x_j \in (d_{(i-1)j}, d_{ij}), m = \overline{1, M}, i = \overline{1, t_j}; \\ 0 \text{ в противоположном случае.} \end{cases}$$

Число приведенных к бинарному виду (бинаризованных) признаков априорного словаря $T_a = \sum_{j=1}^{N_a} t_j$, где t_j — число уровней (интервалов кодирования) признака X_j . Число строк T_p сокращенной БРМ, содержащих элементы x_{ij} , равно $\sum_{j=1}^{N_p} t_j$, где N_p — число признаков рабочего словаря, $N_p \leq N_a$. Записи в БРМ упорядочены по убыванию разделяющей способности бинаризованных признаков в соответствии с их рангом R .

Возможность легкой семантической интерпретации решающего правила (РП) обеспечивается процессом обучения, который сводится к выставлению пороговых уровней и единичных элементов матрицы. При распознавании достаточно выполнить логические операции конъюнкции и импликации в сокращенном пространстве бинарных признаков рабочего словаря X_{T_p} .

Интерпретация «белого ящика»

В отличие от таблицы и дерева решений, конструируемых по описаниям разделяемых подмножеств объектов через диапазоны признаков, БРМ настраивается через интервалы кодирования признаков как характеристические свойства, указывающие на номера классов объектов.



■ **Рис. 1.** Параметрическая идентификация решающего правила
 ■ **Fig. 1.** Parametric identification of the solving rule

Интерпретация «белого ящика» осуществляется в процессе ввода пользователем когнитивной системы значений признаков с помощью универсального алгоритма обработки БРМ (рис. 2). Предъявляемый объект распознается в дискретизированном при обучении пространстве. Процедура распознавания сводится к поэлементной конъюнкции тех строк матрицы, которые определяются атрибутами объекта. Выделенный единственный элемент указывает номер класса.

Решающее правило осуществляет выбор класса объектов на основе комбинации установленных значений бинаризованных признаков $(X_1, X_2, \dots, X_N) \xrightarrow{\text{РП}(X_1, X_2, \dots, X_N)} \Omega_m$, N — глубина поиска, $N \leq N_p$. Наличие единого алгоритма преобразования пространства признаков большой размерности в пространство решений позволяет выполнить визуализацию класса в форме когнитивного образа.

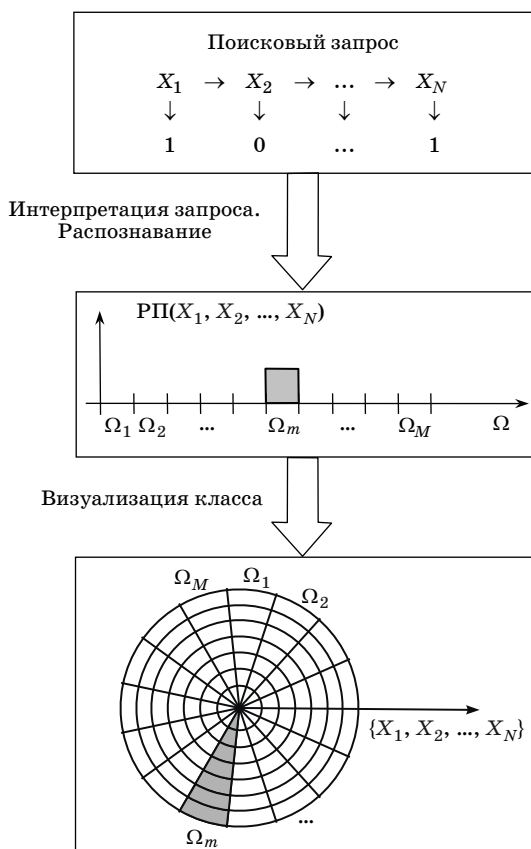
Целью разработки метода визуализации является однозначное и точное отображение класса объекта и возможность отображения характерных и общих признаков классов [18, 19].

Метод визуализации решения с использованием когнитивного облака точек может выпол-

няться после преобразования признакового пространства в пространство главных компонент [20] либо после проецирования данных на заданную пользователем плоскость. Однако многомерность данных требует значительных усилий по поиску оптимальной ориентации плоскости проекции.

Для целостного восприятия когнитивного образа вводимые пользователем значения признаков следует интегрировать в простую геометрическую фигуру, такую как круг или правильный многоугольник. Представление кластерной структуры данных и сокращение объема описаний объектов должны базироваться на образе «мишень» в виде круговой диаграммы, разбитой на сектора, и «калейдоскоп» [21].

Предлагаемый метод визуализации класса выполняет построение когнитивной круговой диаграммы. Ячейка диаграммы определяет отношение между комбинацией значений текущего подмножества признаков $\{X_1, X_2, \dots, X_j\}$ и классом объектов Ω_m . Инициализированные сегменты диаграммы указывают множество возможных решений. Для итогового решения выводятся характеристики выбранного класса.



■ **Рис. 2.** Распознавание и визуализация класса
 ■ **Fig. 2.** Recognition and visualization of a class

Когнитивная диаграмма-визуализатор классов

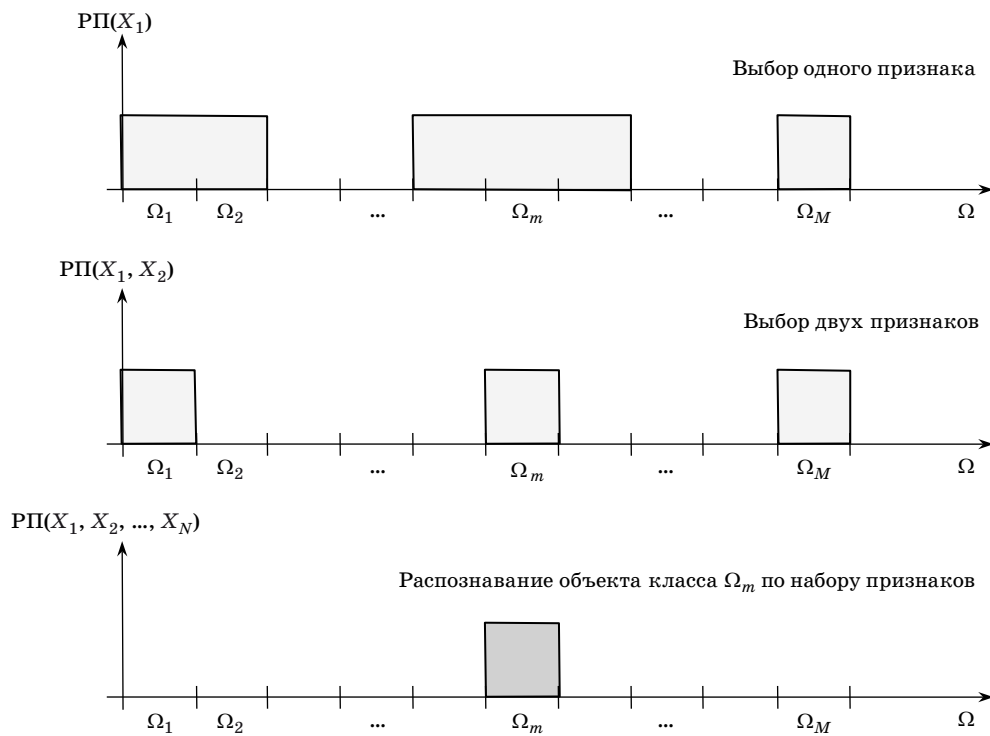
Использование визуализатора классов является заключительным этапом при выборе шаблона, прототипа решения с помощью когнитивной системы анализа данных. Для быстрого поиска объекта характерного типа формируется соответствующая последовательность запросов на ввод пользователем значений признаков.

При задании интервала кодирования очередного признака с наибольшей разделяющей силой из решающей матрицы выбирается строка с указателями на номера классов объектов. Ее поэлементная конъюнкция с установленными ранее ячейками сужает подмножество классов — потенциальных претендентов. Таким образом, допустимое множество решений последовательно уточняется через набор указателей классов, значения признаков которых соответствуют выбранному сочетанию интервалов (рис. 3).

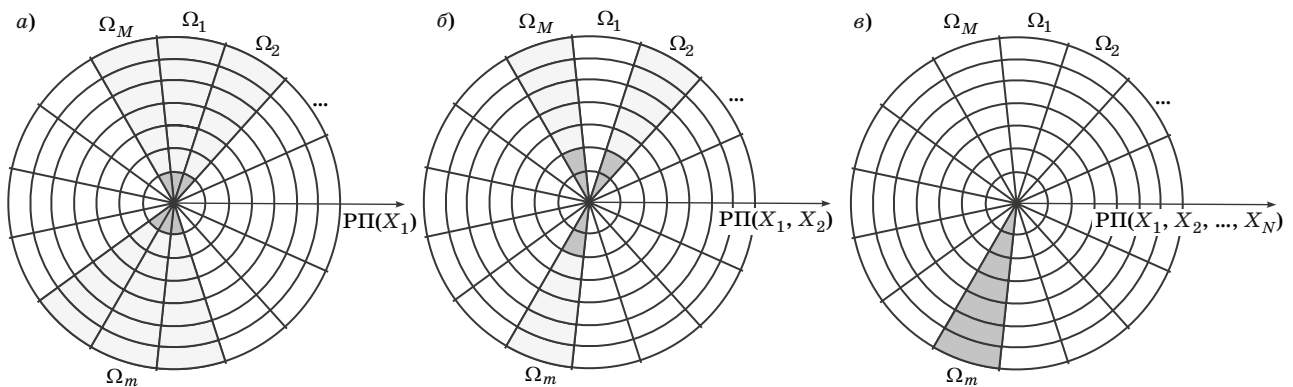
Если на N -м шаге в кодовой строке установлен один элемент, то его положение указывает номер выделенного класса. При упорядоченном рабочем словаре

$$\mathbf{X}_p = \{X_1, \dots, X_j, \dots, X_{T_p} \mid I(X_1) \geq I(X_2) \geq \dots \geq I(X_j) \geq \dots \geq I(X_{T_p})\},$$

где $T_p = |\mathbf{X}_p|$; $I(X_j)$ — дискриминирующая сила X_j [22–23], решение находится без перебора значе-



■ **Рис. 3.** Распознавание объекта по набору бинаризованных признаков
 ■ **Fig. 3.** Object recognition by a set of binarized features



■ **Рис. 4.** Визуализатор распознавания: а — выбор одного признака; б — выбор двух признаков; в — распознавание объекта класса Ω_m по набору признаков
 ■ **Fig. 4.** Recognition visualizer: а — single feature selection; б — choice of two features; в — recognition of a class Ω_m object by a set of features

ний всех признаков. В большинстве случаев распознавание типа объекта останавливается при глубине просмотренного пространства признаков 7–8 % [24].

Визуализатор распознавания демонстрирует на каждом шаге инициированные сегменты диаграммы и соответствующее сужение в подмножестве классифицирующих признаков пространства поиска «подмножество классов» (рис. 4):

$$X_1 \overline{\text{РП}(X_1)} \{ \Omega_1, \Omega_2, \Omega_{m-1}, \Omega_m, \Omega_{m+1}, \Omega_M \};$$

$$(X_1, X_2) \overline{\text{РП}(X_1, X_2)} \{ \Omega_2, \Omega_m, \Omega_M \};$$

$$(X_1, X_2, \dots, X_N) \overline{\text{РП}(X_1, X_2, \dots, X_N)} \Omega_m.$$

Размерность просмотренного пространства определяется кольцами мерности [25]. Азимут инициированной ячейки диаграммы, имеющей наибольшую размерность, указывает принадлежность объекта с установленными признаками целевому классу.

Иницированной ячейке соответствует продукционное правило:

если $((x_1 \in (d_{(i-1)1}, d_{i1})_m$ и ... $x_j \in (d_{(i-1)j}, d_{ij})_m$
и ... $x_N \in (d_{(i-1)N}, d_{iN})_m)$, то $\omega \in \Omega_m$,

где x_j — значение j -го признака, $j = \overline{1, N}$; $(d_{(i-1)j}, d_{ij})_m$ — i -й интервал значений для объектов класса Ω_m .

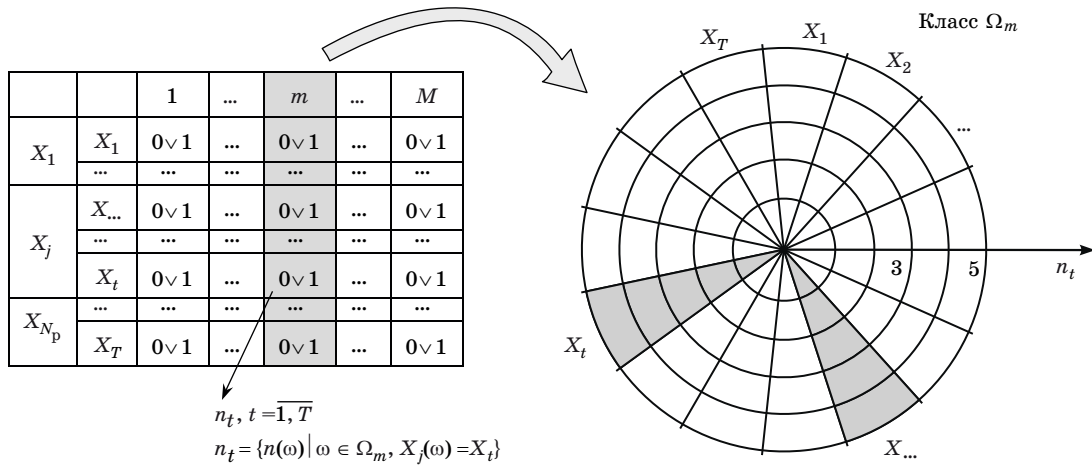
Дальнейшая интерпретация когнитивной диаграммы выполняется в терминах исходных признаков. Совокупность названий интервальных (номинальных) признаков и их значения раскрывают в установленном кластере морфологическое описание типового объекта. Его атрибуты являются наиболее частыми значениями признаков объектов в кластере и определяются как величина моды M_o .

Также указываются количество образцов в найденном кластере и их характерные особенности, отражающие индивидуальность кластера. Для этого рассматривается фрагмент матрицы с объектами целевого кластера.

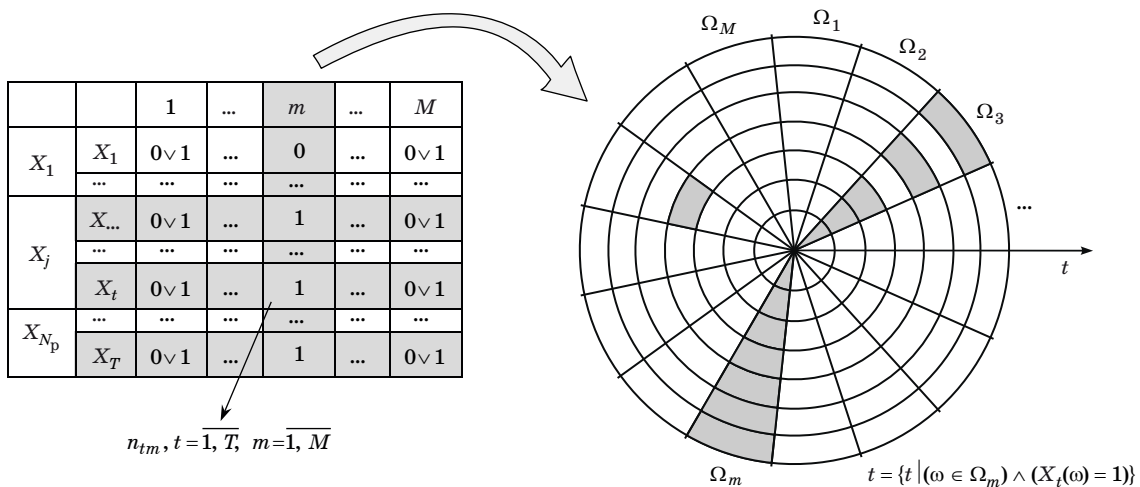
Визуализация характерных и общих признаков

При интерпретации выделенного класса важно указать его отличительные свойства, присущие большинству объектов класса, и те признаки, которые указывают на общие свойства классов.

Характерные особенности объектов класса Ω_m определяются на основе запроса к ячейкам БРМ t -го столбца (рис. 5). В них подсчитывается число объектов обучающей выборки n_t , для которых



■ **Рис. 5.** Визуализатор «Характерные особенности объектов класса Ω_m »
 ■ **Fig. 5.** Visualizer “Characteristic features of class Ω_m objects”



■ **Рис. 6.** Визуализатор «Общие признаки базового класса Ω_m и остальных классов»
 ■ **Fig. 6.** Visualizer “General features of base class Ω_m and other classes”

признак X_j попал в t -й интервал, $t = \overline{1, T}$. На диаграмме число объектов n_t оценивается кольцами мерности. Инициализация сегментов диаграммы происходит при превышении n_t порогового значения $n_{\text{пор}}$, равного 4...6. Из диаграммы следует, что для класса Ω_m характерными являются значения признака $X_j = x_{\dots}$ и $X_j = x_t$, $j = \overline{1, N_p}$.

Интерпретация значений исходных признаков, характерных для большинства объектов класса, выполняется на основе запроса к ячейкам тех строк БРМ, которые соответствуют набору установленных бинарных признаков X_t базового класса Ω_m (рис. 6). В каждом классе определяется число тех объектов n_{tm} , у которых также имеется бинарный признак базового класса.

На диаграмме кольцами мерности оценивается номер t установленного бинарного признака базового класса. Инициализация ячеек диаграммы происходит при превышении n_{tm} порогового значения $n_{\text{пор}}$. Значение $n_{\text{пор}}$ подбирается таким образом, чтобы максимально исключить сегменты, заполненные инициализированными ячейками менее чем на 10 %.

Диаграмма демонстрирует наличие признаков X_t объектов класса Ω_m у других классов. Общие, межклассовые свойства X_1, X_2, X_3, X_{T-2} и X_T наблюдаются у объектов класса Ω_m и класса Ω_3 .

Заключение

Развитие информационного подхода как основы информационного моделирования связано с использованием данных и внедрением методов решения прикладных задач, управляемых данными, когда обученная на выборке система вы-

полняет процедуру распознавания вместо выполнения детерминированного алгоритма.

Хорошие результаты, которые показывают методы глубокого машинного обучения в задачах технического зрения, обработки текста и распознавания речи, получены благодаря иерархической организации признаков, с использованием многочисленных нелинейных преобразований признаков более низкого уровня в признаки более высокого уровня. Как и автоматически построенные на больших данных деревья решений, такие методы моделируют промежуточные и высокоуровневые абстрактные категории, которые сложно интерпретировать.

Особенность сокращенной БРМ, которая обеспечивает представление в явном виде для каждой градации признака бинарной кодовой комбинации классов, позволяет алгоритмизировать интерпретацию принимаемого решения и визуализировать найденный образ. Выбранный вариант визуализации в наибольшей степени адаптирован к данной особенности, отличающей матрицу от таблицы решений. Предложенные в работе модель машинного обучения в виде сокращенной БРМ и диаграмма-визуализатор многовариантного решения просты в понимании и интерпретации. Построенная модель «белого ящика» объясняет выбранный вариант при помощи обычных продукционных правил. Диаграммы-визуализаторы признаков предоставляют сведения об отличительных свойствах класса, присущих большинству объектов класса, и о тех признаках, которые указывают на общие свойства классов. Во многих случаях распознавание типа объекта останавливается при существенно меньшей глубине просмотренного пространства признаков, чем при полном переборе.

Литература

1. Биктимиров М. Р., Елизаров А. М., Щербаков А. Ю. Тенденции развития технологий обработки больших данных и инструментария хранения разноразмерных данных и аналитики. *Электронные библиотеки*, 2016, т. 19, № 5, с. 390–407. ojs.kpfu.ru/index.php/elbib/article/download/207/60 (дата обращения: 16.04.2019).
2. Подберезкин А. И. и др. *Стратегическое прогнозирование и планирование внешней и оборонной политики: монография. Т. 1: Теоретические основы системы анализа, прогноза и планирования внешней и оборонной политики*. М., МГИМО–Университет, 2015. 796 с.
3. Мамедова Г. А., Зейналова Л. А., Меликова Р. Т. Технологии больших данных в электронном образовании. *Открытое образование*, 2017, т. 21, № 6, с. 41–48. doi:http://dx.doi.org/10.21686/1818-4243-2017-6-41-48

4. Болбаков Р. Г. Большие данные в информационных науках. *Образовательные ресурсы и технологии*, 2017, № 1 (18), с. 30–35. doi:10.21777/2312-5500-2017-1-30-35
5. Пименов И. В., Пименов В. И. Методы анализа данных и приобретения знаний при проектировании многомерных объектов дизайна. *Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 1. Естественные и технические науки*, 2017, № 2, с. 3–7.
6. Воронов М. В. Модель технологического действия. *Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 1. Естественные и технические науки*, 2016, № 1, с. 47–52.
7. Одинцова М. А., Радчикова Н. П. Разработка методики самоактивации личности. *Психологические исследования*, 2018, т. 11, № 58. http://psystudy.ru/index.php/num/2018v11n58/1558-odintsova58.html (дата обращения: 16.04.2019).

8. Загоруйко Н. Г., Борисова И. А., Кутненко О. А., Дюбанов В. В., Леванов Д. А. Конкурентное сходство как универсальный базовый инструмент когнитивного анализа данных. *Онтология проектирования*, 2015, т. 5, № 1(15), с. 7–18.
9. Мусаев А. А., Нозик А. А., Русинов Л. А. Прогностический анализ безопасности промышленного предприятия. *Известия Санкт-Петербургского государственного технологического института (технического университета)*, 2016, № 34(60), с. 87–93. doi:10.15217/issn1998984-9.2016.34.87
10. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Networks*, 2015, vol. 61, pp. 85–117. doi:10.1016/j.neunet.2014.09.003
11. Xuan L., Xiaoguang W., Stan M. Improving the interpretability of deep neural networks with knowledge distillation. *IEEE International Conference on Data Mining Workshops (ICDMW)*, 2018, pp. 905–912. doi:10.1109/IJCNN.2018.8489172
12. Богачев И. В., Левенец А. В., Чье Ен Ун. Применение искусственной нейронной сети для классификации телеметрических данных в системах сжатия. *Информационно-управляющие системы*, 2016, № 3, с. 2–7. doi:10.15217/issn1684-8853.2016.3.2
13. Jankowski D., Jackowski K., Cyganek B. Learning decision trees from data streams with concept drift. *Procedia Computer Science*, 2016, vol. 80, pp. 1682–1691. doi:10.1016/j.procs.2016.05.508
14. Etinger D., Simić S.D., Buljubašić L. Automated decision-making with DMN: from decision trees to decision tables. *42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2019, pp. 1514–1518. doi:10.23919/MIPRO.2019.8756694
15. Rong Geng, Xiaojie Wang, Ning Ye, Jun Liu. A fault prediction algorithm based on rough sets and back propagation neural network for vehicular networks. *IEEE Access*, 2018, vol. 6, pp. 74984–74992. doi:10.1109/ACCESS.2018.2881890
16. Пименов И. В., Макаров А. Г. Многомерный анализ и обработка данных при извлечении знаний в области дизайна. *Информатизация образования и науки*, 2015, № 4(28), с. 83–96.
17. Фридман А. Я. Управление комбинированным выводом в дискретных интеллектуальных системах. *Информационно-управляющие системы*, 2015, № 3, с. 35–39. doi:10.15217/issn1684-8853.2015.3.35
18. Ying Zhao, Feng Luo, Minghui Chen, Yingchao Wang, Jiazhi Xia, Fangfang Zhou, Yunhai Wang, Yi Chen, Wei Chen. Evaluating multi-dimensional visualizations for understanding fuzzy clusters. *IEEE Transactions on Visualization and Computer Graphics*, 2019, vol. 25, iss. 1, pp. 12–21. doi:10.1109/TVCG.2018.2865020
19. Annie Preston, Kwan-Liu Ma. Cluster-based visualization for merger tree data: The challenge of missing expectations. *IEEE Scientific Visualization Conference (SciVis)*, 2018, October 21–26, pp. 42–47. doi:10.1109/SciVis.2018.8823586
20. Лазутин О. Г. Методика доведения информации о техническом состоянии космических средств с использованием алгоритмов сжатия данных и когнитивного графического представления. *Труды Военно-космической академии им. А. Ф. Можайского*, 2016, вып. 650, с. 11–17.
21. Емельянова Ю. Г., Фраленко В. П. Методы когнитивно-графического представления информации для эффективного мониторинга сложных технических систем. *Программные системы: теория и приложения*, 2018, т. 9, № 4(39), с. 117–158.
22. Носов А. Л. Системы искусственного интеллекта принятия оптимальных решений в экономике. *Концепт*, 2015, № 11, с. 1–5. <http://e-koncept.ru/2015/15375.htm> (дата обращения: 17.04.2019).
23. Хименко В. И., Охтилев М. Ю., Ключарев А. А., Матьяш В. А. Анализ информативных признаков в задачах обработки данных аэрокосмического мониторинга. *Информационно-управляющие системы*, 2017, № 2, с. 2–12. doi:10.15217/issn1684-8853.2017.2.2
24. Пименов И. В., Ермин Д. А. Автоматизация поиска оптимального дизайн-решения на основе методологии инженерии знаний. *Известия вузов. Технология легкой промышленности*, 2016, № 1, с. 47–50.
25. Staffin T., Flynn B. EVM data visualization: The technomics radar tool (TRT). *ICEAA Professional Development & Training Workshop*, 2019, May 14–17, Tampa, Florida. <http://www.iceaaonline.com/ready/wp-content/uploads/2019/06/SP05-Paper-EVM-Visualization-The-Radar-Tool-Staffin.pdf> (дата обращения: 16.04.2019).

UDC 004.89

doi:10.31799/1684-8853-2019-6-2-11

The cognitive visualization of classifying rules extracted from data based on binary solver matrix modelV. I. Pimenov^a, Dr. Sc., Tech., Professor, orcid.org/0000-0002-7228-3009, v_pim@mail.ruM. V. Voronov^b, Dr. Sc., Tech., Professor, orcid.org/0000-0002-0691-7245I. V. Pimenov^c, PhD, Tech., Associate Professor, orcid.org/0000-0002-1954-6463^aSaint-Petersburg State University of Industrial Technologies and Design, 18, B. Morskaya St., 191186, Saint-Petersburg, Russian Federation^bMoscow State University of Psychology & Education, 24, bld. 27, Otkrytoye Hwy., 107143, Moscow, Russian Federation^cAdmiral Makarov State University of Maritime and Inland Shipping, 5/7, Dvinskaya St., 198035, Saint-Petersburg, Russian Federation

Introduction: Huge volumes of data are generated in cyberspace or from internal information of various organizations. In order to obtain a set of output data with a clear structure, divide it into significant parts and develop rules of classification, machine learning methods are used. Most inductive methods simulate intermediate and high-level abstract categories in multidimensional space which are difficult to interpret. **Purpose:** Developing a model of machine learning in the form of a “white box” which explains the chosen solution using conventional production rules, along with cognitive visualizers for characterizing classes of objects. **Methods:** Formation of a binary decision matrix containing information about a combination of the selected informative sign values which imply the specified classes. **Results:** A binary decision matrix is formed automatically according to the results of cluster and discriminant analyzes. The learning procedure is reduced to setting interval thresholds and matrix elements, which makes it easy to implement a semantic interpretation of a solving rule. The object is recognized by elementwise conjunction of the matrix cells to which the values of the attributes are pointing, and by selection of a single cell corresponding to the class code. To interpret a rule, a universal algorithm for processing a binary matrix has been developed, which applies user-entered attribute values. The dimension of the viewed space is specified by adjustment rings on the recognition visualizer. The azimuth of an initiated diagram cell with the greatest dimensionality indicates the belonging of an object with the set features to a target class. For the characterization of classes, visualizers have been developed, demonstrating both the distinctive properties of a class and properties that several classes share. In many cases, the object type recognition stops when the depth of the scanned features space is significantly less than with a full search. **Practical relevance:** The proposed methods of cognitive analysis and data visualization provide not only the classification of data, determination of the significance of features, their ranking and selection, but also the development of rules which reveal the cause-and-effect relationship between the combination of factors and the type of a made decision.

Keywords — big data, cognitive technologies, multivariate data analysis, machine learning, solving rule, binary decision matrix, visualizer.

For citation: Pimenov V. I., Voronov M. V., Pimenov I. V. The cognitive visualization of classifying rules extracted from data based on binary solver matrix model. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2019, no. 6, pp. 2–11 (In Russian). doi:10.31799/1684-8853-2019-6-2-11

References

1. Biktimirov M. R., Elizarov A. M., Scherbakov A. Yu. Technology trends handling of big data and tools storage of multiformat data and analytics. *Russian Digital Libraries Journal*, 2016, vol. 19, no. 5, pp. 390–407. Available at: ojs.kpfu.ru/index.php/elbib/article/download/207/60 (accessed 16 April 2019) (In Russian).
2. Podberezkin A. I., et al. *Strategicheskoe prognozirovanie i planirovanie vneshnej i oboronnoj politiki: monografiya. T. 1: Teoreticheskie osnovy sistemy analiza, prognoza i planirovaniya vneshnej i oboronnoj politiki* [Strategic forecasting and planning of foreign and defense policy: monograph. Vol. 1. Theoretical foundations of a system for analyzing, forecasting, and planning foreign and defense policy]. Moscow, MGIMO–Universitet Publ., 2015. 796 p. (In Russian).
3. Mamedova G. A., Zeynalova L. A., Melikova R. T. Big data technologies in e-learning. *Open Education*, 2017, vol. 21, no. 6, pp. 41–48 (In Russian). doi:http://dx.doi.org/10.21686/1818-4243-2017-6-41-48
4. Bolbakov R. G. Big data in information sciences. *Obrazovatel'nye resursy i tekhnologii*, 2017, no. 1 (18), pp. 30–35 (In Russian). doi:10.21777/2312-5500-2017-1-30-35
5. Pimenov I. V., Pimenov V. I. Methods of data analysis and knowledge extraction at the projecting of multivariate design objects. *Vestnik Sankt-Peterburgskogo gosudarstvennogo universiteta tekhnologii i dizajna. Seriya 1. Estestvennye i tekhnicheskie nauki*, 2017, no. 2, pp. 3–7 (In Russian).
6. Voronov M. V. Modeling of technological processes. *Vestnik Sankt-Peterburgskogo gosudarstvennogo universiteta tekhnologii i dizajna. Seriya 1. Estestvennye i tekhnicheskie nauki*, 2016, no. 1, pp. 47–52 (In Russian).
7. Odintsova M. A., Radchikova N. P. The development of personal self-activation inventory. *Psikhologicheskie issledovaniya*, 2018, vol. 11, no. 58. Available at: http://psystudy.ru/index.php/num/2018v11n58/1558-odintsova58.html (accessed 16 April 2019) (In Russian).
8. Zagoruiko N. G., Borisova I. A., Kutenko O. A., Dyubanov V. V., Levanov D. A. Rival similarity as an universal basic tool of cognitive data mining. *Ontology of Designing*, 2015, vol. 5, no. 1(15), pp. 7–18 (In Russian).
9. Musaev A. A., Nozick A. A., Rusinov L. A. Predictive analysis of safety of an industrial enterprise. *Bulletin of the Saint-Petersburg State Institute of Technology (Technical University)*, 2016, no. 34(60), pp. 87–93 (In Russian). doi:10.15217/issn1998984-9.2016.34.87
10. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Networks*, 2015, vol. 61, pp. 85–117. doi:10.1016/j.neunet.2014.09.003
11. Xuan L., Xiaoguang W., Stan M. Improving the interpretability of deep neural networks with knowledge distillation. *IEEE International Conference on Data Mining Workshops (ICDMW)*, 2018, pp. 905–912. doi:10.1109/IJCNN.2018.8489172
12. Bogachev I. V., Levenets A. V., Chye En Un. Artificial neural networks in classifying telemetry data for compression systems. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2016, no. 3, pp. 2–7 (In Russian). doi:10.15217/issn1684-8853.2016.3.2
13. Jankowski D., Jackowski K., Cyganek B. Learning decision trees from data streams with concept drift. *Procedia Computer Science*, 2016, vol. 80, pp. 1682–1691. doi:10.1016/j.procs.2016.05.508
14. Etinger D., Simić S. D., Buljubašić L. Automated decision-making with DMN: from decision trees to decision tables. *42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2019, pp. 1514–1518. doi:10.23919/MIPRO.2019.8756694
15. Rong Geng, Xiaojie Wang, Ning Ye, Jun Liu. A fault prediction algorithm based on rough sets and back propagation neural network for vehicular networks. *IEEE Access*, 2018, vol. 6, pp. 74984–74992. doi:10.1109/ACCESS.2018.2881890

16. Pimenov I. V., Makarov A. G. The multidimensional analysis and data processing at knowledge extraction in the field of design. *Informatizaciya obrazovaniya i nauki*, 2015, no. 4(28), pp. 83–96 (In Russian).
17. Fridman A. Ya. Control of combined inference in discrete intelligence systems. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2015, no. 3, pp. 35–39 (In Russian). doi:10.15217/issn1684-8853.2015.3.35
18. Ying Zhao, Feng Luo, Minghui Chen, Yingchao Wang, Jiazhi Xia, Fangfang Zhou, Yunhai Wang, Yi Chen, Wei Chen. Evaluating multi-dimensional visualizations for understanding fuzzy clusters. *IEEE Transactions on Visualization and Computer Graphics*, 2019, vol. 25, iss. 1, pp. 12–21. doi:10.1109/TVCG.2018.2865020
19. Annie Preston, Kwan-Liu Ma. Cluster-based visualization for merger tree data: The challenge of missing expectations. *IEEE Scientific Visualization Conference (SciVis)*, 2018, October 21–26, pp. 42–47. doi:0.1109/SciVis.2018.8823586
20. Lazutin O. G. Technique of communicating information about the technical state of space vehicles using data compression algorithms and cognitive graphical representation. *Trudy Voenno-kosmicheskoy akademii imeni A. F. Mozhajskogo*, 2016, vol. 650, pp. 11–17 (In Russian).
21. Emelyanova Ju. G., Fralenko V. P. Methods of cognitive-graphical representation of information for effective monitoring of complex technical systems. *Program Systems: Theory and Applications*, 2018, vol. 9, no. 4(39), pp. 117–158 (In Russian).
22. Nosov A. L. Artificial intelligence systems for optimal decision making in the economy. *Koncept*, 2015, no. 11, pp. 1–5. Available at: <http://e-koncept.ru/2015/15375.htm> (accessed 17 April 2019) (In Russian).
23. Khimenko V. I., Okhtilev M. Yu., Klucharev A. A., Matish V. A. Informative feature analysis in data processing for aerospace monitoring. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2017, no. 2, pp. 2–12 (In Russian). doi:10.15217/issn1684-8853.2017.2.2
24. Pimenov I. V., Ermin D. A. Automation of optimum design decision search on the basis of knowledge engineering methodology. *The News of Higher Educational Institutions. Technology of Light Industry*, 2016, no. 1, pp. 47–50 (In Russian).
25. Staffin T., Flynn B. EVM data visualization: The technomics radar tool (TRT). *ICEAA Professional Development & Training Workshop*, 2019, May 14–17, Tampa, Florida. Available at: <http://www.iceaaonline.com/ready/wp-content/uploads/2019/06/SP05-Paper-EVM-Visualization-The-Radar-Tool-Staffin.pdf> (accessed 16 April 2019).

УВАЖАЕМЫЕ АВТОРЫ!

Научная электронная библиотека (НЭБ) продолжает работу по реализации проекта SCIENCE INDEX. После того как Вы зарегистрируетесь на сайте НЭБ (<http://elibrary.ru/defaultx.asp>), будет создана Ваша личная страничка, содержание которой составят не только Ваши персональные данные, но и перечень всех Ваших печатных трудов, имеющих в базе данных НЭБ, включая диссертации, патенты и тезисы к конференциям, а также сравнительные индексы цитирования: РИНЦ (Российский индекс научного цитирования), h (индекс Хирша) от Web of Science и h от Scopus. После создания базового варианта Вашей персональной страницы Вы получите код доступа, который позволит Вам редактировать информацию, помогая создавать максимально объективную картину Вашей научной активности и цитирования Ваших трудов.