

На правах рукописи



ПИМЕНОВ Илья Викторович

**МЕТОДЫ И АЛГОРИТМЫ ИЗВЛЕЧЕНИЯ ЗНАНИЙ ДЛЯ
ИНТЕЛЛЕКТУАЛЬНОГО ПОИСКА ДИЗАЙНЕРСКОГО РЕШЕНИЯ**

Специальность 05.13.01 – Системный анализ, управление и обработка
информации (технические системы)

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

Санкт-Петербург – 2017

Работа выполнена в федеральном государственном бюджетном образовательном учреждении высшего образования “Санкт-Петербургский государственный университет промышленных технологий и дизайна” (СПбГУПТД)

Научный руководитель: Доктор технических наук, профессор,
Макаров Авинир Геннадьевич
СПбГУПТД,
заведующий кафедрой интеллектуальных
систем и защиты информации

Официальные оппоненты: Доктор технических наук, профессор,
Ходаковский Валентин Аветикович
Федеральное государственное бюджетное
образовательное учреждение высшего об-
разования «Петербургский государствен-
ный университет путей сообщения Импе-
ратора Александра I»
заведующий кафедрой «Математика и мо-
делирование»

кандидат технических наук
Толмачев Сергей Геннадьевич
АО «Концерн «Гранит-Электрон»,
начальник НИЛ-5032

Ведущая организация: Санкт-Петербургский государственный
технологический институт (технический
университет)

Защита состоится «05» октября 2017 г. в __.00 часов на заседании диссертационного совета Д.002.199.01 при Федеральном государственном бюджетном учреждении науки Санкт-Петербургском институте информатики и автоматизации Российской академии наук по адресу: 199178, Санкт-Петербург, В.О., 14 линия, 39.

С диссертацией можно ознакомиться в библиотеке Федерального государственного бюджетного учреждения науки Санкт-Петербургского института информатики и автоматизации Российской академии наук, <http://www.spiiras.nw.ru/ru>.

Автореферат разослан “__” _____ 201__ года.

Ученый секретарь
диссертационного совета Д 002.199.01
доктор технических наук

С.В. Кулешов

Общая характеристика работы

Актуальность темы диссертации. Дизайн как важный инструмент конкурентной борьбы широко используется в индустрии моды, при разработке промышленной продукции, мобильных устройств, программного обеспечения. Система, представляющая процесс дизайна, должна обеспечивать функциональность, стиль, технологичность и другие формальные качества объекта, который имеет множество первичных характеристик, описывающих образ, назначение, конструкцию, эстетическую ценность, эргономичность и другие свойства. Качество многомерного объекта определяется не только его морфологической структурой, но и скрытым, неявным взаимодействием между формообразующими частями. Выявление законов, объясняющих, как значения первичных характеристик и их сочетания влияют на показатели дизайна, необходимо при подготовке специалистов, проектировании новых изделий на основе установленных прототипов и их отличительных свойств.

В настоящее время накоплены значительные массивы данных “объект–свойства” в различных областях дизайна. Их использование позволяет объективизировать оценку дизайн-решения, опираясь на значения измеримых признаков найденного прототипа. Инструментальной базой при извлечении из данных знаний для широкой совокупности объектов-прецедентов являются методы многомерного анализа данных и машинного обучения. Их комплексное использование является сложной системной задачей ввиду разнотипности признаков, наличия взаимосвязей между ними, необходимостью организации данных при передаче между методами. Результат в виде алгоритма распознавания определяет принадлежность объекта к одному из альтернативных классов посредством геометрического представления. Проблемой остается дальнейший переход к продукционным правилам, обеспечивающим семантическую интерпретацию решений в условиях многомерности описаний объектов, когда число свойств объектов дизайна варьируется от ста до нескольких сотен, и формирование базы знаний при построении интеллектуальной системы.

Поэтому актуальной является разработка комплексного подхода, основанного на применении многомерного анализа данных, методов и алгоритмов для построения баз знаний, позволяющих повысить степень автоматизации создания интеллектуальных систем в области дизайна.

Степень разработанности темы. Развитию методологических основ системного подхода к исследованиям в области дизайн-проектирования способствовали труды российских и зарубежных ученых: А.Н. Родченко, Б. Н. Гусева, Е. Я. Сурженко, Е. Б. Кобляковой, А. С. Далидович, Г. П. Щедровицкого, М. А. Коськова, А. Н. Лаврентьева, Б. Арчера, А. И. Половинкина. Вопросы разработки интеллектуальных САПР с поддержкой 3D-моделирования одежды рассмотрены в работах К. А. Процик, И. Л. Ключко, О. А. Мишенина. Вопросы дизайна пользовательских веб-интерфейсов рассмотрены в работах Д. В. Кирсанова, Д. К. Сатина, Д. В. Бородаева, Г. П. Блуднова, П. Макнейла, Дж. Гарретта, Я. Нильсена, А. Купера.

Процесс дизайн-проектирования, связанный с применением конструкторских приемов, гармонизацией композиционных и цветовых решений, является трудноформализуемой задачей, что оправдывает разработку систем, основанных на знаниях в области дизайна. Развитию методов проектирования интеллектуальных систем и баз

знаний (БЗ) в значительной степени способствовали труды В. В. Андрейчикова, А. И. Башмакова, Т. А. Гавриловой, В. И. Городецкого, В. Ф. Хорошевского, Т. Б. Чистяковой, В. А. Дюка, Д. В. Сошникова, Е. Е. Витяева, Х. Уэно, Д. Элти и др.

Цель и задачи исследования. Целью диссертационной работы является повышение степени автоматизации создания интеллектуальных систем в области дизайна на основе комплексного подхода к использованию многомерного анализа данных, методов и алгоритмов для построения баз знаний и интеллектуального поиска дизайнерского решения. Для достижения указанной цели в работе решены следующие задачи:

1) анализ современного состояния задачи формализации поиска дизайн-решения и построения модели знаний об объекте дизайна, выявление методов многомерного анализа данных, составляющих инструментальную базу для создания интеллектуальных систем;

2) разработка комплексного подхода к использованию многомерного анализа данных для построения распознающих баз знаний и интеллектуальных систем в области дизайна;

3) разработка методов отбора и ранжировки признаков по их вкладу в разделяющую силу, а также выбора локальных пространств для покрытия классов объектов дизайна;

4) разработка алгоритма формирования логического решающего правила по результатам кластерного и дискриминантного анализов;

5) разработка алгоритма интеллектуального поиска на основе распознающей базы знаний;

6) проверка эффективности предложенных методов и алгоритмов при разработке ряда интеллектуальных систем, основанных на знаниях в области дизайна.

Научная новизна работы:

1) разработан подход к построению баз знаний интеллектуальных систем в области дизайна, отличающийся комплексным применением методов многомерного анализа данных и обеспечивающий извлечение знаний из массива "объекты-свойства", в том числе установление правил, раскрывающих причинно-следственные связи между свойствами объекта дизайна и его типом;

2) разработаны методы двухэтапного отбора, ранжировки признаков и выбора локальных пространств, отличающиеся применением модели множественной пошаговой регрессии и обеспечивающие учет нарушения принципа аддитивности при рассмотрении вклада переменных в их совместную разделяющую силу, автоматизированное построение системы решающих правил, а также снижение трудоемкости при поиске покрытия каждого типа объектов в виде конъюнкции элементарных событий о попадании значений признаков в определенные интервалы;

3) разработан алгоритм формирования логического решающего правила, позволяющий использовать результаты машинного обучения для построения распознающей базы знаний (представлять найденные решающие правила в базе знаний);

4) разработан алгоритм интеллектуального поиска, отличающийся применением распознающей базы знаний, не требующий построения промежуточных понятий для реализации последовательных стратегий поиска дизайн-решения и выполняющий семантическую интерпретацию многомерного объекта.

Практическая значимость. Разработанный в диссертационной работе подход к использованию многомерного анализа данных позволяет автоматизировать процесс

построения баз знаний и повысить эффективность создания интеллектуальных систем в области дизайна.

Предложенный в работе комплекс методов и алгоритмов позволяет выполнять построение баз знаний для ряда трудноформализуемых задач дизайна. Использование распознающей базы знаний, сформированной по результатам обмера населения, позволяет перейти от размерной к морфологической типологии, основанной на внутренних соотношениях между частями проектируемого изделия, для учета индивидуальных особенностей потребителей.

Представленная в интеллектуальной системе структура предметной области позволяет повысить точность расчета среднетипичных размеров изделий при серийном производстве за счет компенсации смещения оценок коэффициентов регрессии, чувствительных к расположению “периферийных” наблюдений для редких типов пользователей, и обеспечивает хорошие прогностические свойства найденных расчетных уравнений для всего диапазона размеров.

Применение разработанных методов и алгоритмов позволяет формировать базы знаний экспертных систем, проектировать интеллектуальные базы данных и поисковые системы в областях, связанных с выявлением прототипа, шаблона, поиском многомерного объекта-образца. Знания об иерархической структуре классов, информативности признаков и правилах описания классов на языке признаков позволяют находить эффективный путь следования запросов, без перебора значений всех признаков, и повысить релевантность результатов поиска в отобранном кластере.

Основные результаты, полученные в работе, доведены до уровня расчетных методик и вычислительных алгоритмов, что облегчает их применение при создании интеллектуальных систем.

Основные методы исследования. Методической и теоретической основой диссертационного исследования явились научные труды по системному анализу, теории распознавания образов. Широко использовались современные методы многомерного анализа данных, искусственного интеллекта.

Компьютерная реализация разработанных алгоритмов производилась на основе объектно-ориентированного подхода.

Положения, выносимые на защиту:

1) подход к построению баз знаний интеллектуальных систем в области дизайна, основанный на комплексном применении методов многомерного анализа данных, обеспечивает извлечение знаний из массива “объекты-свойства” и установление правил, раскрывающих причинно-следственные связи между свойствами объекта дизайна и его типом;

2) методы двухэтапного отбора, ранжировки признаков и выбора локальных пространств обеспечивают учет нарушения принципа аддитивности при рассмотрении вклада переменных в их совместную разделяющую силу, автоматизированное построение системы решающих правил и снижение трудоемкости при поиске покрытия каждого типа объектов в виде конъюнкции элементарных событий о попадании значений признаков в определенные интервалы;

3) алгоритм формирования логического решающего правила позволяет использовать результаты машинного обучения для построения распознающей базы знаний;

4) алгоритм интеллектуального поиска, использующий представленные в базе знаний решающие правила, не требует построения промежуточных понятий для реализации последовательных стратегий поиска дизайн-решения и выполняет семантическую интерпретацию многомерного объекта.

Достоверность полученных результатов обеспечивалась корректностью исходных предпосылок и преобразований при получении статистических зависимостей, а также примерами практической реализации и апробацией основных результатов работы на конференциях и в научной печати.

Реализация результатов работы. Результаты диссертационной работы реализованы, внедрены и используются в виде методики построения баз знаний интеллектуальных систем, методов и алгоритмов интеллектуального анализа данных в ООО "КОЖИНФОРМДИЗАЙН" для проектирования мужских перчаток установленного морфологического типа, расчета среднетипичных размеров перчаток при серийном производстве, в ФГБУН Музее антропологии и этнографии им. Петра Великого (Кунсткамера) РАН для интеллектуального поиска музейного образца – прототипа народного костюма, в МИП "Девега" для извлечения знаний в области веб-дизайна, в Санкт-Петербургском государственном университете промышленных технологий и дизайна, Балтийском государственном техническом университете "ВОЕНМЕХ" им. Д. Ф. Устинова, позволяя сократить затраты на обработку и систематизацию больших массивов данных, объективизировать оценку проектного решения, повысить точность расчета размерных признаков по сравнению с существующими методиками проектирования, что подтверждено соответствующими актами.

Разработанный подход к формализации задач дизайна и построения баз знаний широко используются в учебном процессе СПбГУПТД в курсах "Методы и средства исследований", "Информационное обеспечение дизайн-проектирования", "Информационные системы в дизайне изделий легкой промышленности", "Математическое моделирование", "Интеллектуальные информационные системы" для студентов различных направлений подготовки.

Апробация работы. Основные результаты работы докладывались, одобрены и опубликованы в материалах следующих конференций: международных научно-методических конференциях "Информатизация инженерного образования (Инфорино 2014)" (Москва, 2014), "Совершенствование математического образования – 2014: проблемы и пути их решения" (Тирасполь, 2014); международных конференций "Математическое моделирование в образовании, науке и производстве" (Тирасполь, 2011, 2012, 2013, 2015), "Информатика: проблемы, методология, технологии" (Воронеж, 2013, 2014); конференции "Математика в вузе и школе" (Гатчина, 2012); всероссийской научной конференции "Нейрокомпьютеры и их применение" (Москва, 2014); всероссийских научных конференций молодых ученых "Инновации молодежной науки" (Санкт-Петербург, 2010, 2011, 2013, 2014, 2016).

Публикации

Основные результаты диссертации отражены в 29 печатных работах, в том числе 10 статьях в научных журналах из перечня ВАК РФ, 17 докладах на международных и всероссийских конференциях, 2 свидетельствах о регистрации программ.

Структура и объем работы

Диссертация состоит из введения, четырех разделов, заключения, списка использованных источников из 203 наименований, 5 приложений. Общий объем работы – 204 страницы, в том числе основной текст – 158 страниц, 13 таблиц, 52 рисунка.

Содержание работы

Во введении изложены основные положения диссертационной работы, обоснована актуальность темы, определена цель исследований и решаемые задачи, дана характеристика научной новизны и практической значимости работы. Представлены данные об апробации работы и публикациях.

В первой главе проведен анализ современного состояния задачи поиска дизайн-решения и построения модели знаний об объекте дизайна.

Поиск дизайна сводится к выбору альтернатив, отражающих как внутреннюю структуру объекта, так и его внешний образ. При проектировании одежды осуществляется выбор композиционно-конструктивного решения; при проектировании обуви, перчаток – выбор формы, морфологии изделия; при создании сайта, модели одежды важен выбор стиля.

Для описания конструкции, композиции и морфологии изделия используются измеряемые признаки, характеризующие размеры, форму, пропорции элементов, их цвет. Для описания стиля широко используются экспертные оценки. Классификация композиционно-конструктивных и морфологических решений может выполняться многомерными методами, стиль же изделия определяется экспертом.

Исходные данные об объектах-прецедентах $\omega_i, i = \overline{1, n}$ представляются значениями признаков $(x_{i1}, \dots, x_{ij}, \dots, x_{iN_A})$ априорного словаря X_A , которые характеризуют свойства $\{X_j | j = \overline{1, N_A}\}$ изготовленных объектов, и сведениями о принадлежности объектов-прецедентов к одному из выявленных классов $\Omega_m, \Omega_m \subset \Omega, \Omega = \Omega_1 \cup \Omega_2 \cup \dots \cup \Omega_M$, которые описываются классификационной переменной Y .

Для извлечения знаний из данных (Data Mining) используются две основные группы методов: индуктивные методы машинного обучения и статистические методы. При этом создается модель заданной структуры (например, нейронная сеть, дискриминантные функции, линейная регрессия), а ее параметры подстраиваются под данные об объектах.

В экспертных системах знания специалистов эксплицируются с помощью коммуникативных и текстологических методов. Ансамбли моделей усложняют интерпретацию результатов.

Наиболее удобными, с точки зрения семантической интерпретации, являются логические решающие правила (РП), которые используют по каждому признаку правила соответствия между двумя объектами ОВ ω_l и ω_q , определяемые интервалами порогового кодирования $[x_{l_0j}, x_{l_kj}]$:

$$d_j(\omega_q, \omega_l) = \begin{cases} 1, & \text{если } x_{l_0j} \leq x_{qj} \leq x_{l_kj}, \\ 0 & \text{в противоположном случае,} \end{cases}$$

ω_l – объект m -го класса, характеризуемый вектором $(x_j)_l = (x_{l1}, \dots, x_{lj}, \dots, x_{lN_A})$, $l = \overline{1, |\Omega_m|}$, $m = \overline{1, M}$.

Покрытие m -го класса логическими классификаторами D_{sm} выглядит как:

$$D_{sm}(\omega_q, \omega_l) = \bigwedge_{j=1}^{N_A} d_j(\omega_q, \omega_l) \bigwedge \bigvee_{r=1}^{\overline{n_m}} \bigwedge_{j=1}^{N_A} d_j(\omega_r, \omega_l),$$

где $s_m = \overline{1, S_m}$, S_m – число классификаторов m -го класса; $r = \overline{1, n_m}$, $\overline{n_m} = n - n_m = |\Omega \setminus \Omega_m|$, $n_m = |\Omega_m|$ – число объектов ОВ m -го класса; $\omega_r \in \Omega \setminus \Omega_m$; $\bigcup_{m=1}^M \Omega_m = \Omega$.

Продукционное правило относит объект ω к тому классу, в чей классификатор D_{s_m} он попал, то есть $\omega \in \Omega_m$, если $\bigvee_{s_m} D_{s_m}(\omega, \omega_m \in \Omega_j) = 1$.

Инструментальной базой при извлечении из данных знаний, для широкой совокупности объектов-прецедентов, являются рассмотренные в разделе методы многомерного анализа данных. Поскольку число свойств объектов дизайна приближается к сотне, осуществить умозрительно работу по систематизации правил, описывающих динамическую функциональную структуру модели знаний, представляется невозможным. Целенаправленное применение многомерного анализа для структуризации понятий, представляющих объекты дизайна, и построения методом распознавания образов решающих правил, раскрывающих причинно-следственные связи между свойствами изделия и его показателями качества, повышение доли автоматических процедур в процессе организации и передачи данных между методами многомерного анализа являются основой предложенного подхода к построению интеллектуальных систем в области дизайна.

Вторая глава посвящена разработке комплексного подхода к использованию многомерного анализа данных для построения баз знаний, содержащих неявные правила и закономерности в области дизайна.

Исходный массив “объекты–свойства” исследуется в глобальном пространстве измеримых свойств объекта дизайна. Для обеспечения численного представления качественных свойств номинальные и порядковые переменные преобразуют к бинарному типу.

Сжатие признакового пространства выполняется в два этапа (рис. 1). Предварительный отбор признаков $X_{ГК} \cup X_{КЛ} \cup X_{ДА}$ осуществляется с помощью методов главных компонент, кластеризации переменных и дискриминантного анализа. Вычисления проводятся по формулам для количественных признаков с использованием евклидовой метрики.

Для кластеризации объектов используется метод Варда, основанный на дисперсионном анализе для оценки расстояний между кластерами. Расстояние между кластерами Ω_{m1} и Ω_{m2} вычисляется по принципу “средней связи”, как среднее арифметическое всех возможных пар комбинаций между объектами ω_i и ω_j , входящими в различные кластеры:

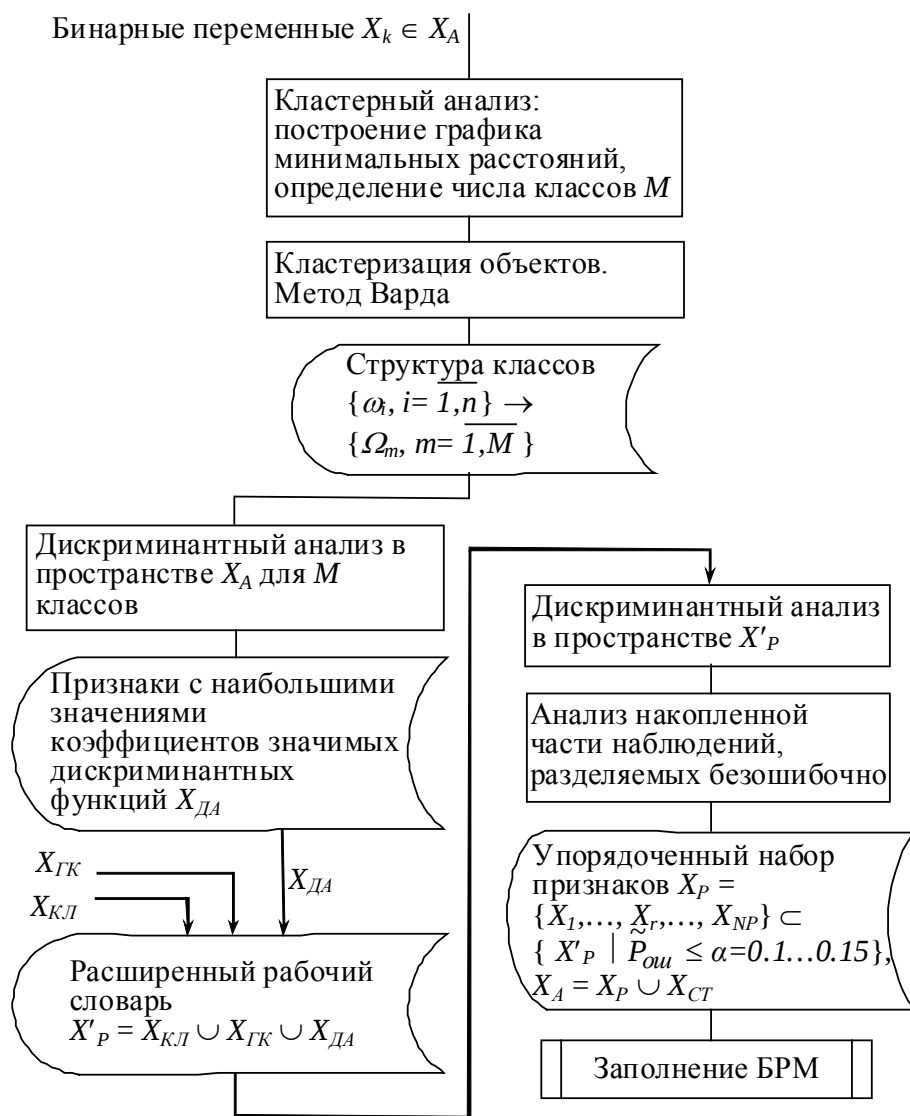
$$d(\Omega_{m1}, \Omega_{m2}) = \frac{1}{n_{m1} n_{m2}} \sum_{\omega_i \in \Omega_{m1}} \sum_{\omega_j \in \Omega_{m2}} d(\omega_i, \omega_j).$$

Измерение расстояний при наличии периферийных выбросов в пространстве главных компонент выполняется с помощью метрики “манхэттенское расстояние” (City Block):

$$d(\omega_i, \omega_k) = \sum_{j=1}^{N_A} |x_{ij} - x_{kj}|.$$



a



б

Рис.1. Схема многомерного анализа:

a – кластеризация переменных и использование метода главных компонент;

б – использование кластерного и дискриминантного анализов

Ввиду того, что дискриминирующая сила набора признаков не следует принципу аддитивности, на втором этапе выполняется ранжировка признаков по росту накопленной части безошибочных разделений объектов, как показано на схеме справа. В рабочем словаре остаются признаки, достаточные для разделения 90–95% объектов выборки.

Дискриминантные функции не могут в явном виде использоваться для построения баз знаний, поскольку не содержат сведений о порядке индуктивного логического вывода. Поэтому формируется бинарная решающая матрица (БРМ), которая может автоматически обрабатываться в процессе логического вывода или транслироваться в поле знаний.

Третья глава посвящена разработке алгоритмов формирования решающего правила и интеллектуального поиска на основе построенного правила (рис. 2).

Логическое решающее правило формируется по результатам кластерного и дискриминантного анализов (рис. 2, а). Правило, в форме бинарной матрицы, осуществляет разбиение пространства признаков на области с помощью гиперплоскостей, ортогональных координатным осям, и легко интерпретируется.

В нем содержатся сведения о сочетании значений отобранных информативных признаков предъявляемого объекта и принадлежности его к заданному классу.

Параметрическая идентификация правила заключается в установке единичных значений ячеек матрицы x_{ij} при наличии бинарного признака у объектов m -го класса, либо в случае принадлежности значений количественного признака X_j интервалу с номером t

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

Простота процедуры обучения, которая сводится к установке интервальных порогов и элементов матрицы, позволяет легко реализовать семантическую интерпретацию решающего правила. При распознавании достаточно выполнить логические операции и операции сравнения.

Алгоритм интеллектуального поиска основан на обработке решающего правила (рис. 2, б). Он реализует процедурную часть базы знаний интеллектуальной системы для выбора прототипа дизайн-объекта.

В процессе интеллектуального поиска в ячейках рабочей памяти устанавливаются единичные значения, указывающие на претендентов для дальнейшего рассмотрения. Множество индексов классов последовательно уточняется по выбранному пользователем значению или интервалу признака.

Таким образом, база знаний, сформированная в процессе обработки результатов многомерного анализа данных с использованием автоматизированных процедур, включает знания как декларативного, так и процедурного типов. Декларативная часть содержит знания о структуре предметной области, морфологии объектов и разделительной силе их признаков. Процедурная же часть реализуется как универсальный алгоритм обработки БРМ, использующий вводимые пользователем интеллектуальной системы значения признаков.

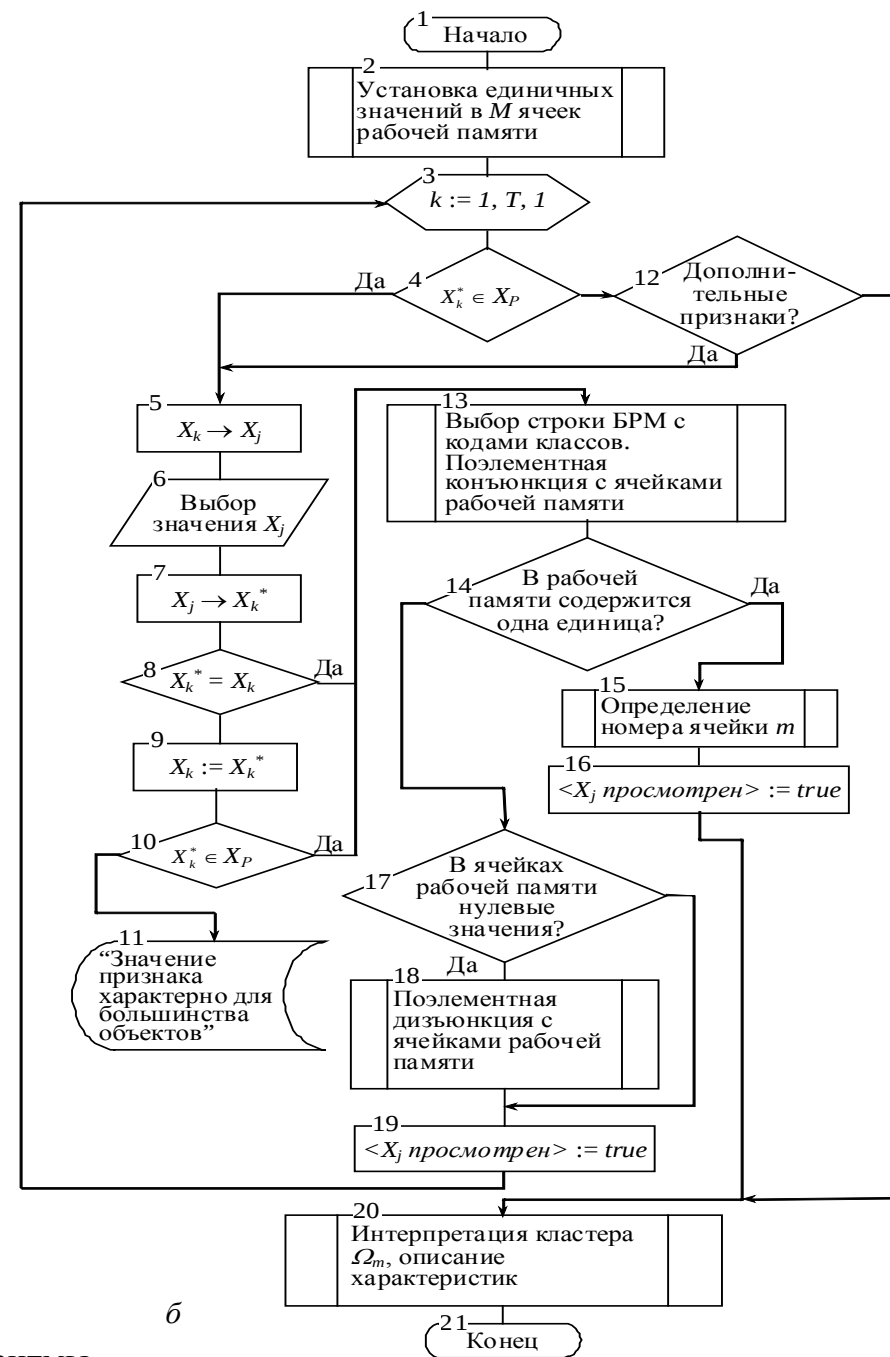
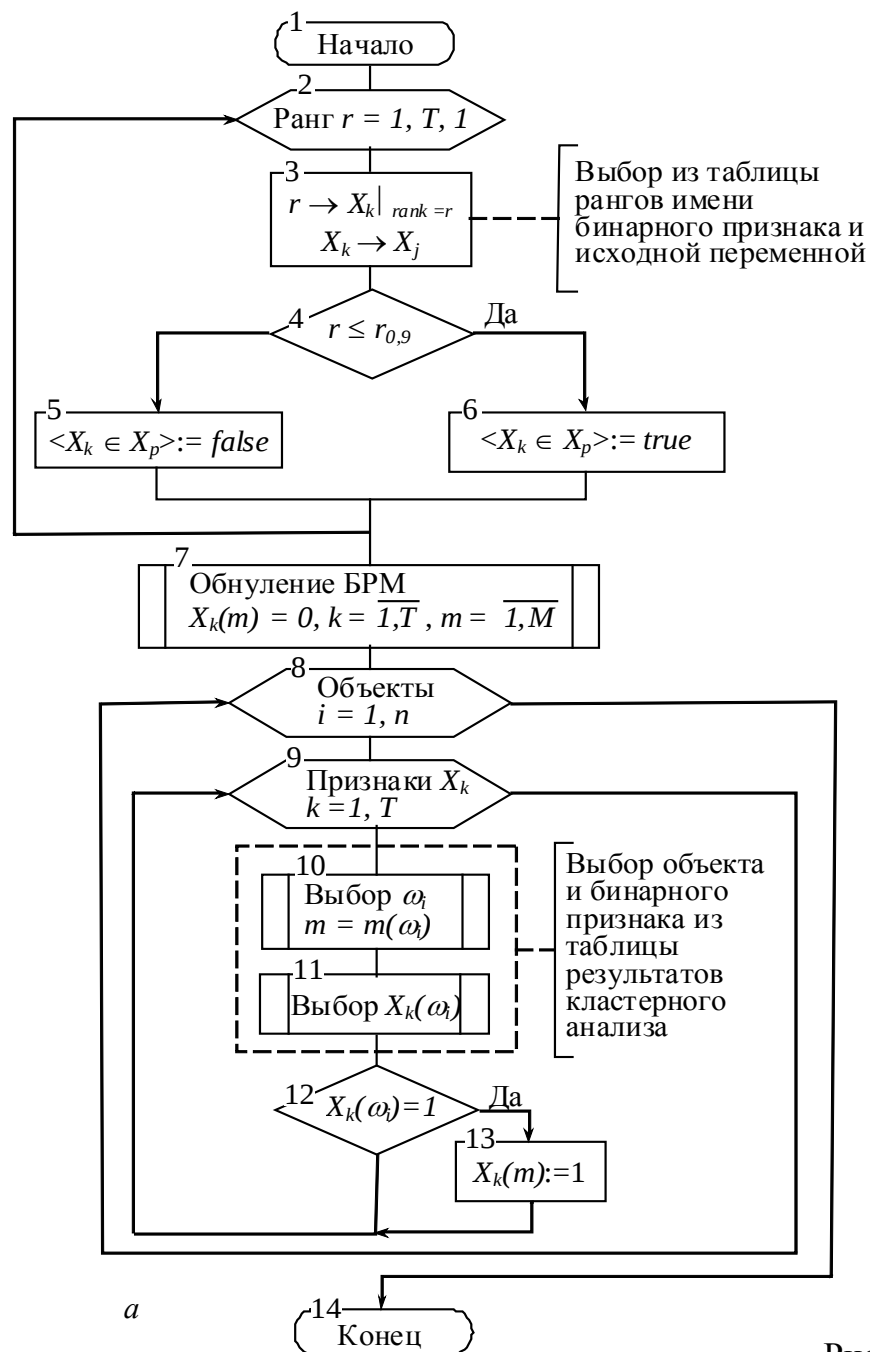


Рис.2. Алгоритмы:

а – формирование решающего правила; б – интеллектуальный поиск на основе решающего правила

В четвертой главе решена задача проверки эффективности предложенных методов и алгоритмов при разработке трех интеллектуальных систем. Выбор композиционно-конструктивных решений рассмотрен на примере музейных образцов традиционных женских рубаш (ТЖР, количество исходных признаков $N_A = 98$, объем выборки $n = 200$, количество основных кластеров $M = 6$), морфологических решений – при проектировании мужских перчаток ($N_A = 77$, $n = 503$, $M = 7$), стилистических решений – при описании дизайна веб-сайтов ($N_A = 92$, $n = 200$, $M = 13$).

Вначале подгруппа признаков выбирается по составу первой и второй главных компонент. Они объясняют большую часть изменчивости данных. В подгруппу $X_{ГК}$ отобраны признаки с наибольшим весом.

Затем выполняется кластеризация признаков для выделения однородных групп. Число групп определяется по графику минимальных расстояний, на которых происходит объединение переменных в набор до значительного скачка по шкале дистанций. Состав этих групп $X_{КЛ}$ определяется после построения сосульчатой диаграммы (Icicle plot).

При кластеризации объектов – ТЖР выявлено 6 основных типов рубаш (рис. 3).

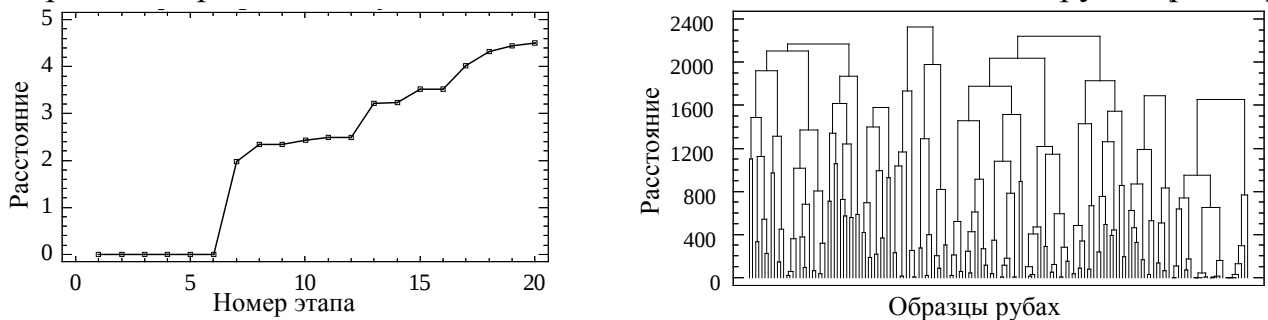


Рис. 3. Кластеризация объектов для ТЖР:

а – график группировки близких объектов (агломеративная процедура, манхэттенское среднее расстояние между кластерами); б – дендограмма (метод Варда, манхэттенское расстояние)

Пять дискриминантных функций позволяют правильно разделить 98,8% наблюдений (рис. 4). По наибольшим значениям их стандартизированных коэффициентов отобрано 12 признаков в состав подгруппы $X_{ДА}$.

На первом этапе отобрана совокупность $X_{КЛ} \cup X_{ГК} \cup X_{ДА}$ из 24 признаков. Далее в пространстве отобранных признаков проведен дискриминантный анализ. С помощью накопленной части наблюдений, разделяемых безошибочно, определены ранги признаков, дающих наибольший вклад в разделение классов. В результате двухэтапной процедуры из 98 исходных признаков отобран рабочий словарь X_P , включающий 15 признаков – соответствующие дискриминантные функции безошибочно разделяют более 90% образцов.

Преобразование данных в процессе обучения и использования интеллектуальной системы представлено на рис.5. Результатами машинного обучения являются:

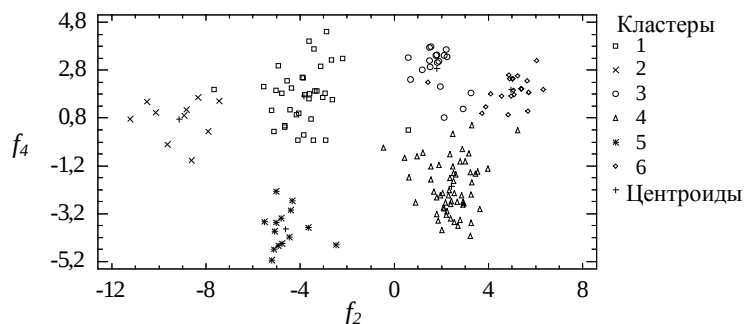


Рис. 4. Дискриминантный анализ ТЖР. Диаграмма рассеяния объектов на плоскости функций f_2, f_4

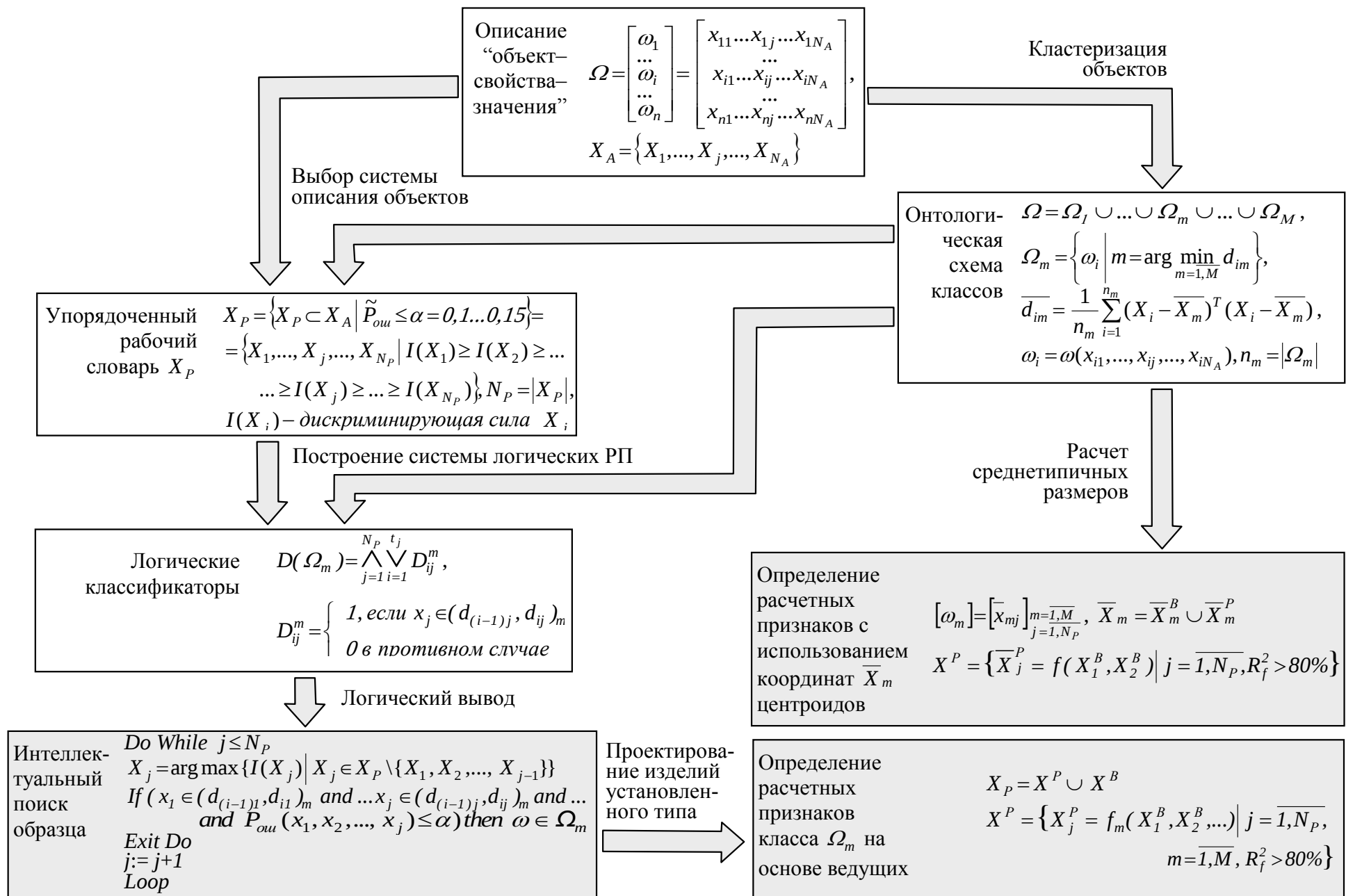


Рис.5. Преобразование данных в процессе обучения и использования интеллектуальной системы

онтологическая схема распределения объектов ω_i по классам Ω_m , рабочий словарь признаков X_P , упорядоченный по их дискриминирующей силе $I(X_j)$, РП в виде логических классификаторов $D(\Omega_m)$, описывающих кластеры в терминах признаков и устанавливающих правила D_{ij}^m соответствия i -го интервала j -го признака классу Ω_m . В процессе логического вывода задается преобразование значений признаков в классы объектов:

$$\text{if } ((x_1 \in (d_{(i-1)1}, d_{i1})_m \text{ and } \dots x_j \in (d_{(i-1)j}, d_{ij})_m \text{ and } \dots x_N \in (d_{(i-1)N}, d_{iN})_m) \\ \text{then } \omega \in \Omega_m,$$

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

При поиске образца пользователю предъявляются признаки X_j , обладающие максимальной дискриминирующей силой среди еще не рассмотренных. Решение принимается, когда оценка вероятности ошибки распознавания объектов $\tilde{P}_{oui}$ не будет превышать заданного уровня α , соответствующего 10–15%. Решающие правила позволяют значительно сократить глубину поиска объекта-прототипа. Выбор музейного образца ТЖР останавливается, в большинстве случаев, после просмотра 6–8 признаков. Для найденного объекта выводится изображение и дается композиционно-конструктивное описание – названия признаков и их значения.

Использование полученных знаний при проектировании изделий установленного типа дает возможность установить морфологический тип потребителя и сформировать адекватные зависимости между размерными и ведущими признаками изделий. Для каждого типа строятся собственные взаимосвязи, учитывающие особенности кластера.

В условиях серийного производства используются среднетипичные размеры. Поиск зависимостей по координатам центроидов кластеров позволяет учесть результат кластеризации и повысить качество моделей регрессии. Например, высоким коэффициентом детерминации R^2 обладают найденные уравнения регрессии для расчетного значения длины кисти с ладонной стороны D_lad и ширины кисти с тыльной стороны Wk_tyl , используемые при проектировании перчаток:

$$D_lad = 92,48 + 0,47 \cdot O_kist_kost, R^2 = 80\%,$$

$$Wk_tyl = -9,5 + 0,48 \cdot O_kist_kost, R^2 = 93\%,$$

где основной признак O_kist_kost – обхват кисти на уровне головки пястной кости. Найденные же в рамках традиционной методики уравнения имеют низкое качество – построенное по всей выборке линейное уравнение для D_lad характеризуется значением R^2 , равным 27%.

Определение в локальном пространстве дискриминантных функций на основе модели пошаговой регрессии использовалось при описании области веб-дизайна. Для отделения стилистического класса Ω_m от остальных создается классификационная переменная

$$y_m(x) = \begin{cases} 1, \text{ если } \omega_i(x) \in \Omega_m, \\ -1 \text{ в противоположном случае.} \end{cases}$$

Регрессионная зависимость для этой переменной является границей класса и содержит отличительные признаки класса, описывая семантику исследуемого стиля.

В приложении приведены исходные данные об объектах дизайна и их характеристики – параметры обмера кистей мужских рук, признаки традиционных составных женских рубаш, характеристики дизайна веб-страниц, критерии классификации сайтов; представлены экранные изображения разработанных программ для сбора статистических данных в области веб-дизайна и инструментальной среды для проектирования базы знаний интеллектуальных систем, а также акты внедрения результатов диссертационной работы.

Заключение

Подход, основанный на комплексном применении многомерного анализа данных, совокупность предложенных методов и алгоритмов построения баз знаний продукционного типа, а также их практическая реализация представляют собой решение актуальной научно-технической задачи автоматизации создания интеллектуальных систем для широкой совокупности объектов дизайна, обеспечивая семантическую интерпретацию решений в условиях многомерности описаний объектов. Их внедрение вносит значительный вклад в развитие инструментальной базы для приобретения знаний и оценки дизайн-решений. При решении данной задачи были получены следующие результаты.

1. Подход, основанный на комплексном использовании методов многомерного анализа, обеспечивает, при выработке проектно-художественных решений, извлечение зафиксированных в изготовленных объектах знаний из массива данных “объекты–свойства”. Помимо классификации данных в области дизайна, структуризации понятий, представляющих объект, определения значимости признаков, выполнения их ранжирования и отбора, устанавливаются правила, раскрывающие причинно-следственные связи между свойствами изделия и его показателями качества, внешнего вида.

2. Методы двухэтапного отбора и ранжировки признаков объектов дизайна, и автоматизированного построения в локальном пространстве системы решающих правил на основе модели множественной пошаговой регрессии, разработанные в рамках комплексного подхода. Предварительный отбор признаков осуществляется с помощью анализа главных компонент, кластеризации переменных и дискриминантного анализа. На втором этапе, учитывая нарушение принципа аддитивности при рассмотрении вклада переменных в их совместную разделяющую силу, выполняется ранжировка признаков по росту накопленной части безошибочных разделений объектов. Формирование описаний типов объектов в локальном пространстве наиболее характерных значений признаков значительно снижает трудоемкость при поиске покрытия каждого класса в виде конъюнкции элементарных событий о попадании значений признаков в определенные интервалы.

3. Алгоритм формирования логического решающего правила по результатам кластерного и дискриминантного анализов, осуществляющий представление результатов машинного обучения в виде распознающей базы знаний. Правило выполняет разбиение пространства признаков на области с помощью гиперплоскостей, ортого-

нальных координатным осям, и легко интерпретируется. Параметрическая идентификация правила заключается в установке единичных значений ячеек памяти при наличии бинарного признака у объектов m -го класса, либо в случае принадлежности значений количественного признака рассматриваемому интервалу кодирования.

4. Алгоритм интеллектуального поиска, использующий вводимые пользователем значения признаков, основан на обработке логического решающего правила. Алгоритм реализует процедурную часть базы знаний интеллектуальной системы для выбора шаблона дизайн-решения.

5. Методы, алгоритмы, базы знаний и интеллектуальные системы, разработанные в рамках предложенного подхода, используются для проектирования мужских перчаток установленного морфологического типа, расчета среднетипичных размеров перчаток при серийном производстве, извлечения знаний в области веб-дизайна, интеллектуального поиска музейного образца – прототипа народного костюма, позволяя сократить затраты на обработку и систематизацию больших массивов данных, объективизировать оценку проектного решения, повысить точность расчета размерных признаков по сравнению с существующими методиками проектирования, что подтверждено соответствующими актами внедрения и свидетельствами о государственной регистрации программ для ЭВМ.

Дальнейшая интеграция инструментов машинного обучения и инженерии онтологий позволит выявлять как структуру классов, так и составляющих элементов дизайна, выстраивать цепочки промежуточных понятий, учитывать технологические и конструктивные ограничения, что в конечном итоге послужит основой для поддержки процессов синтеза дизайн-решений на основе накопленных знаний об элементах дизайна и их взаимосвязях. Перспективными направлениями исследования являются также выявление предпочтений потребителя, тенденций моды и прогнозирование стилей, на основе накопленных данных о динамике параметров объектов дизайна.

Полученные результаты соответствуют п. 4. “Разработка методов и алгоритмов решения задач системного анализа, оптимизации, управления, принятия решений и обработки информации”, п. 5. “Разработка специального математического и программного обеспечения систем анализа, оптимизации, управления, принятия решений и обработки информации”, п. 12. “Визуализация, трансформация и анализ информации на основе компьютерных методов обработки информации”, п. 13. “Методы получения, анализа и обработки экспертной информации” паспорта научной специальности 05.13.01 “Системный анализ, управление и обработка информации (технические системы)”.

Список основных публикаций, отражающих содержание работы

Статьи в журналах, входящих в "перечень ВАК..."

1. Пименов, В. И. Совершенствование методики проектирования перчаток на основе интеллектуального анализа данных / В. И. Пименов, В. В. Семенова, И. В. Пименов, А. В. Михеева // Дизайн. Материалы. Технология. – 2012. – № 4(24). – С. 101–106.

2. *Пименов, И. В.* Построение распознающих баз знаний для поиска дизайн-решений / И. В. Пименов // Системы управления и информационные технологии. – 2014. – № 1.1(55). – С. 183–186.
3. *Пименов, И. В.* Разработка методики создания баз знаний в области дизайна / И. В. Пименов, А. Г. Макаров, М. Л. Шатковская, В. И. Пименов // Дизайн. Материалы. Технология. – 2014. – № 4(34). – С. 90–96.
4. *Пименов, И. В.* Использование знаний при проектировании изделий установленного типа / И. В. Пименов, В. В. Семенова, В. И. Пименов // Системы управления и информационные технологии. – 2015. – № 2(60). – С. 86–90.
5. *Пименов, И. В.* Многомерный анализ и обработка данных при извлечении знаний в области дизайна / И. В. Пименов, А. Г. Макаров // Информатизация образования и науки. – 2015. – № 4(28). – С. 83–96.
6. *Пименов, И. В.* Подход к интеллектуальному поиску дизайн-решения / И. В. Пименов, К. В. Григорьева, В. И. Пименов // Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 1. Естественные и технические науки. – 2015. – № 4. – С. 29–35.
7. *Пименов, И. В.* Использование распознающей базы знаний при проектировании объектов дизайна / И. В. Пименов, В. И. Пименов // Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 1. Естественные и технические науки. – 2016. – № 1. – С. 32–36.
8. *Пименов, И. В.* Автоматизация поиска оптимального дизайн-решения на основе методологии инженерии знаний / И. В. Пименов, Д. А. Ермин // Известия вузов. Технология легкой промышленности. – 2016. – № 1. – С. 47–50.
9. *Пименов, И. В.* Использование знаний в информационной системе расчета средне-типичных размеров при серийном производстве / И. В. Пименов, Н. Р. Туркина, В. И. Пименов // Автоматизация. Современные технологии. – 2016. – № 4. – С. 23–27.
10. *Пименов, И. В.* Автоматизация создания интеллектуальной системы на основе методов локальной геометрии / И. В. Пименов, Т. А. Кравец, В. И. Пименов // Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 1. Естественные и технические науки. – 2016. – № 4. – Сер. 1. – С. 9–14.

Статьи, материалы конференций

11. *Пименов, И. В.* Исследовательский комплекс в области web-дизайна / И. В. Пименов // Вестник молодых ученых Санкт-Петербургского государственного университета технологии и дизайна: сб. науч. тр.: в 4 ч. Ч. 1: Естественные и технические науки. – СПб.: СПГУТД, 2011. – С. 201–206.
12. *Пименов, В. И.* Концептуальный анализ области Web-дизайна методами многомерного анализа данных / В. И. Пименов, И. В. Пименов // Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 1. Естественные и технические науки. – 2011. – № 2. – С. 66–71.
13. *Пименов, И. В.* Многомерный анализ данных при проектировании сложных объемных изделий / И. В. Пименов // Труды докладов междунар. научно-методич. конф. “Математика в вузе и в школе”, СПб.: ПГУПС – 2012. – С. 176–177.
14. *Пименов, И. В.* Построение морфологической типологии при серийном производстве изделий / И. В. Пименов // Пути решения проблем совершенствования математического образования: интеграция науки и практики: матер. 7 междунар. науч.-практ. конф. – Тирасполь, 2012. – С. 103.

15. *Пименов, И. В.* Методология построения системы интеллектуального поиска музейного образца / И. В. Пименов, В. И. Пименов // Труды докладов XIII междунар. конф. «Информатика: проблемы, методология, технологии», Воронеж, ВГУ – 2013. – С. 49–53.
16. *Пименов, И. В.* Программный комплекс интеллектуальной поисковой системы, основанной на знаниях в области дизайна / И. В. Пименов // Вестник молодых ученых Санкт-Петербургского государственного университета технологии и дизайна. – 2013. – № 1. – С. 47–51.
17. *Пименов, И. В.* Алгоритмизация построения базы знаний материальных и информационных продуктов / И. В. Пименов // Труды докладов междунар. науч.-методич. конф. «Информатизация инженерного образования (Инфорино 2014)», Москва, Издательство МЭИ – 2014. – С. 115–116.
18. *Пименов, В. И.* Методы формализации процесса дизайна / В. И. Пименов, И. В. Пименов // Труды докладов XIV междунар. конф. «Информатика: проблемы, методология, технологии», Воронеж, ВГУ – 2014. – С. 149–152.
19. *Пименов, И. В.* Обработка информации при построении интеллектуальной базы данных / И. В. Пименов, А. Г. Макаров // Труды докладов XIV междунар. конф. «Информатика: проблемы, методология, технологии», Воронеж, ВГУ – 2014. – С. 144–148.
20. *Пименов, И. В.* Интеллектуальная система анализа дизайн-процессов / И. В. Пименов // Труды докладов XII Всерос. науч. конф. «Нейрокомпьютеры и их применение», Москва, МГППУ – 2014. – С. 25.
21. *Пименов, И. В.* Программно-аппаратный комплекс для автоматизации разработки интеллектуальных информационных систем / И. В. Пименов // Вестник молодых ученых Санкт-Петербургского государственного университета технологии и дизайна. – 2014. – № 1. – С. 60–63.
22. *Пименов, В. И.* Инструментальная база когнитивных технологий / В. И. Пименов, А. Г. Макаров, И. В. Пименов // Совершенствование математического образования – 2014: проблемы и пути их решения: материалы VIII междунар. науч.-метод. конф. – Тирасполь, 2014. – С. 207–208.
23. *Пименов, И. В.* Построение системы интеллектуального поиска объекта дизайна / И. В. Пименов // Вестник молодых ученых Санкт-Петербургского государственного университета технологии и дизайна. – 2015. – № 1. – С. 60–64.
24. *Пименов, И. В.* Использование методов многомерного анализа для построения интеллектуальных систем / И. В. Пименов // Вестник молодых ученых Санкт-Петербургского государственного университета промышленных технологий и дизайна. – 2016. – № 2. – С. 24–28.

Свидетельства на регистрацию интеллектуальной собственности

25. *Пименов, И. В.* Свидетельство об официальной регистрации программы для ЭВМ № 2014615514, РФ. Инструментальная среда для проектирования базы знаний интеллектуальных систем / И. В. Пименов, В. И. Пименов, А. Г. Макаров, М. Л. Шатковская; правообладатель СПГУТД. – Зарег. 28.05.2014. – Бюл. № 6. – 1 с.
26. *Пименов, И. В.* Свидетельство об официальной регистрации программы для ЭВМ № 2014618361, РФ. Оценка показателей классификации, критериев качества и характеристик дизайна при анализе сайтов / И. В. Пименов, В. И. Пименов; правообладатель СПГУТД. – Зарег. 18.08.2014. – Бюл. № 9. – 1 с.

Подписано в печать _____.2017. Печать трафаретная
Усл. печ. л. 1,0. Формат 60×84 1/16. Тираж 100 экз. Заказ № ____
Отпечатано в типографии СПбГУПТД
191028, Санкт-Петербург, ул. Моховая, 26