

На правах рукописи

Смагин Сергей Владимирович

**МЕТОД ОЦЕНКИ ЗНАЧЕНИЙ
ПАРАМЕТРОВ МАТЕМАТИЧЕСКОЙ МОДЕЛИ
МЕДИЦИНСКОЙ ДИАГНОСТИКИ**

05.13.18 – Математическое моделирование,
численные методы и комплексы программ

Автореферат

диссертации на соискание ученой степени
кандидата технических наук



Владивосток – 2012

Работа выполнена в лаборатории интеллектуальных систем Института автоматизации и процессов управления Дальневосточного отделения РАН.

Научный руководитель: Клещев Александр Сергеевич
доктор физико-математических наук, профессор

Научный консультант: Черняховская Мэри Юзефовна
доктор медицинских наук

Официальные оппоненты: Девятисильный Александр Сергеевич
доктор технических наук, профессор,
зав. сек. управления и навигации ИАПУ ДВО РАН

Кулешов Евгений Львович
доктор технических наук, профессор,
зав. каф. компьютерных систем ДВФУ

Ведущая организация: Институт математики им. С.Л. Соболева
Сибирского отделения РАН, г. Новосибирск

Защита состоится «11» октября 2012 года в 10:00 часов на заседании диссертационного совета Д.005.007.01 в Институте автоматизации и процессов управления ДВО РАН по адресу: 690041, г. Владивосток, ул. Радио, 5.

С диссертацией можно ознакомиться в библиотеке Института автоматизации и процессов управления Дальневосточного отделения РАН.

Автореферат разослан «___» _____ 2012 г.

Ученый секретарь
диссертационного совета Д.005.007.01
к.т.н.



А.В. Лебедев

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы. Индуктивное формирование знаний на основе эмпирических данных является основным способом получения новых эмпирических знаний в науке и практике. Он заключается в получении общего знания о некоторой совокупности объектов на основании анализа единообразного описания конечного множества отдельных представителей этой совокупности – обучающей выборки данных (или просто обучающей выборки).

Моделирование такого способа познания лежит в основе многих направлений исследований, получивших в англоязычной литературе названия: Data Mining (интеллектуальный анализ данных), Machine Learning (машинное обучение), Knowledge Discovery in Databases (обнаружение знаний в базах данных, сокращенно KDD), Pattern Recognition (распознавание образов), Knowledge Extraction (извлечение знаний), Information Discovery (обнаружение информации), Information Harvesting (сбор информации), Data Archaeology (археология данных) и т.д., каждое из которых характеризуется собственным подходом к проблеме индуктивного формирования знаний, собственными постановками задач и многообразием методов их решения. Значительный вклад в эти направления внесли отечественные и зарубежные ученые: Айзерман М.А., Бонгард М.М., Браверманн Э.М., Вагин В.Н., Вапник В.Н., Витяев Е.Е., Журавлев Ю.И., Загоруйко Н.Г., Клещев А.С., Лбов Г.С., Розоноэр Л.И., Рудаков К.В., Рязанов В.В., Финн В.К., Червоненкис А.Я., Михальский Р.С., Пятецкий-Шапиро Г.И. и другие.

В основополагающих публикациях по данной тематике сформулированы общие постановки основных задач индуктивного формирования знаний – классификации и кластеризации, изучены разнообразные модели зависимости между классами и объектами, а также разработано большое число алгоритмов обучения (классификации и кластеризации), решающих поставленные задачи на этих моделях. Задача классификации состоит в том, чтобы для некоторого множества моделей зависимости, к которому относится неизвестная зависимость между классами и объектами, разработать алгоритм классификации, который на основе описания объектов обучающей выборки строит решающее правило, вероятность правильной классификации которого любых новых объектов как можно выше. Задача кластеризации состоит в том, чтобы, используя некоторую метрику, разработать алгоритм кластеризации, который на основе описания объектов обучающей выборки разбивает обучающую выборку на непересекающиеся подмножества, называемые кластерами, так, чтобы каждый кластер состоял из схожих объектов, а объекты разных кластеров существенно отличались, а также строит описания кластеров, позволяющие относить к ним новые объекты.

Общепризнанной характеристикой алгоритмов классификации считается время обучения – время, за которое алгоритм классификации по обучающей выборке строит решающее правило. Другой характеристикой алгоритмов классификации принято считать качество формируемых с их помощью решающих правил. Основными показателями качества решающих правил считаются вероятность правильной классификации, устойчивость результатов, переобучение и некоторые другие. Для экспериментального исследования показателей качества решающих правил созданы репозитории баз данных, включающие в себя выборки

реальных данных из различных предметных областей, а также разработано несколько универсальных программ генерации модельных данных. Полученные таким образом значения показателей качества для различных решающих правил (на основе одних и тех же выборок) сравниваются между собой, из этого делается вывод о том, какой из алгоритмов классификации лучше.

По степени того, насколько интерпретируемым для эксперта предметной области получается описание классов, формирующие это описание алгоритмы обучения (классификации и кластеризации) могут быть разделены на следующие группы. Численные алгоритмы обучения строят описание классов, обладающее лишь минимальной интерпретируемостью (названия классов в таком описании соответствуют названиям реальных классов предметной области, названия признаков – названиям признаков объектов этой предметной области, а названия значений признаков – реальным значениям признаков). Логические алгоритмы обучения строят описание классов в виде логической функции, которая может быть задана, например, таблицей истинности, которая может интерпретироваться экспертами предметной области. Описание классов, представленное в форме продукций (а также описания, построенные логико-вероятностными алгоритмами обучения и ДСМ-методом), обладает большей интерпретируемостью, чем в предыдущем случае. Описание классов, построенное для моделей зависимости с параметрами, являющихся реальными онтологиями предметных областей (которые получены в результате формализации представлений экспертов о предметных областях, т.е. взяты из практики), представленных небогатыми системами логических соотношений с параметрами, обладает самой высокой степенью интерпретируемости, также как и сами эти модели зависимости.

Однако, в литературе не рассматривается специфическая постановка задачи классификации для моделей зависимости с параметрами, которая требует от алгоритма классификации формирования описаний классов, обладающих определенным уровнем качества. Степень интерпретируемости описаний классов, которые формируют существующие алгоритмы обучения (классификации и кластеризации) для практически полезных предметных областей, не позволяет экспертам предметных областей использовать эти описания в своей практической деятельности. Поэтому актуальной проблемой является разработка алгоритмов обучения (классификации и кластеризации) для практически полезных и хорошо интерпретируемых математических моделей зависимости с параметрами, являющихся реальными онтологиями предметных областей, формирующих такие описания классов (базы знаний), которые эксперты предметных областей оценивают как достаточные для решения практических задач в этих предметных областях.

Целью диссертационной работы является разработка алгоритма классификации для математической модели зависимости с параметрами (являющейся онтологией медицинской диагностики, приближенной к реальной) который по обучающей выборке находит значения параметров (базу знаний) этой модели, близкие к значениям, характеризующим эту предметную область.

Для достижения этой цели необходимо решить следующие задачи:

1. Разработать общую постановку задачи классификации для математических моделей зависимости с параметрами, как частный случай задачи оценки значений параметров модели.

2. С помощью компьютерных экспериментов исследовать вопрос о связи между разработанной постановкой задачи и постановкой задачи классификации, рассматриваемой в литературе.
3. Разработать алгоритм классификации для практически полезной и хорошо интерпретируемой математической модели зависимости с параметрами, являющейся онтологией медицинской диагностики, приближенной к реальной (заданной системой логических соотношений с параметрами), и провести экспериментальное исследование его свойств.
4. Разработать комплекс программ и технологию его применения для индуктивного формирования баз медицинских знаний в форме, принятой в медицинской литературе.

Методы исследования. Для решения указанных задач использовались методы искусственного интеллекта, машинного обучения, распознавания образов, интеллектуального анализа данных, компьютерного моделирования, теории вероятностей и математической статистики, теории множеств, структурного программирования.

Научная новизна работы состоит в следующем:

- разработана общая постановка задачи классификации для математических моделей зависимости с параметрами, как частный случай задачи оценки значений параметров модели (при этом критерием качества обучения является близость оценок значений параметров к значениям, характеризующим предметную область, а не вероятность правильной классификации решающих правил, получаемых в результате обучения);
- показано, что решение задачи классификации в разработанной общей постановке является и решением задачи классификации в традиционной постановке, однако обратное не всегда верно;
- введены новые понятия внешних и внутренних свойств алгоритма классификации для моделей зависимости с параметрами (как функции вероятности определенных событий, связанных с алгоритмом классификации, зависящие от объема обучающих выборок); эти свойства характеризуют алгоритмы классификации и не зависят от конкретных обучающих выборок; разработан также способ оценки значений этих свойств;
- разработан алгоритм классификации для практически полезной, хорошо интерпретируемой и адекватной математической модели зависимости с параметрами, являющейся онтологией медицинской диагностики, приближенной к реальной.

Практическая ценность работы состоит в разработке комплекса программ и технологии его использования, позволяющего создавать обучающие выборки (состоящие из историй болезни различных разделов медицины) и на их основе индуктивно формировать базы медицинских знаний (в форме, принятой в медицинской литературе), а также объяснение этих баз знаний.

Реализация результатов работы. Разработанный комплекс программ был использован для индуктивного формирования базы знаний о заболеваниях раздела медицины «острый живот» на основе обучающей выборки объемом 69 историй болезни. Формальное представление базы знаний позволяет использовать ее

в экспертных системах медицинской диагностики. Высокая степень интерпретируемости и достоверность сформированной базы знаний подтверждены экспертом. Результаты работы использованы в Дальневосточном федеральном университете – в учебном процессе при чтении курса лекций по дисциплине «Системы искусственного интеллекта» студентам специальности 010503.65 «Математическое обеспечение и администрирование информационных систем», а также в научной работе и исследованиях сотрудников лаборатории интеллектуальных систем Института автоматизации и процессов управления ДВО РАН.

Положения, выносимые на защиту: общие постановки задач классификации и кластеризации для математических моделей зависимости с параметрами, представленные как частный случай задачи оценки значений параметров модели; определения внешних и внутренних свойств алгоритмов классификации и кластеризации для моделей зависимости с параметрами и метод оценки значений этих свойств; алгоритм классификации для практически полезной, хорошо интерпретируемой и адекватной математической модели зависимости с параметрами, являющейся онтологией медицинской диагностики, приближенной к реальной; комплекс программ и технология его применения для индуктивного формирования баз медицинских знаний в форме, принятой в медицинской литературе.

Обоснованность и достоверность результатов работы обеспечиваются применением методов искусственного интеллекта и подтверждены результатами компьютерных экспериментов.

Апробация работы. Основные научные и практические результаты работы докладывались и обсуждались на следующих международных и российских конференциях и семинарах: Дальневосточных математических школах-семинарах имени академика Е.В. Золотова (Владивосток: 2004, 2007, 2008), Втором Международном форуме медицинских информационных технологий для разработчиков и пользователей «MedSoft» (Москва: 2006), Седьмой Международной научно-технической конференции «Интеллектуальные и многопроцессорные системы» (Таганрог: 2006), Научных сессиях МИФИ (Москва: 2007, 2008), секции «Вычислительная математика и кибернетика» XVI Международной научной конференции студентов, аспирантов и молодых ученых (Москва: 2009), Всероссийских конференциях с международным участием «Знания – Онтологии – Теории» (Новосибирск: 2009, 2011), конкурсах работ молодых ученых Института автоматизации и процессов управления ДВО РАН (Владивосток: 2009, 2011), семинарах лаборатории интеллектуальных систем Института автоматизации и процессов управления ДВО РАН и базовой кафедры программного обеспечения ЭВМ Дальневосточного федерального университета (Владивосток: 2005-2012).

Публикация результатов работы. По материалам диссертации опубликовано 25 печатных работ, в том числе 4 статьи в журналах, входящих в список ВАК РФ, и 7 статей в других российских и зарубежных журналах.

Структура и объем работы. Диссертационная работа состоит из введения, четырех глав, заключения, списка литературы, включающего 103 наименования, глоссария терминов и 15 приложений. Основная часть работы изложена на 133 страницах, содержит 60 рисунков и 4 таблицы.

СОДЕРЖАНИЕ РАБОТЫ

В **первой главе** представлен обзор литературы по проблеме индуктивного формирования знаний.

Во **второй главе** решаются первые две задачи диссертации. Обозначим: x – вектор переменных (признаков); X – многомерное (признаковое) дискретное пространство, координатами которого являются переменные вектора x ; $y = \{y_1, \dots, y_n\}$ – множество классов; математическая модель зависимости $m(x, y, q)$ – система соотношений между вектором переменных x , значениями переменной y и вектором параметров q с областью возможных дискретных значений Q .

Модель зависимости m индуцирует отношение $K \subset X \times y$ между объектами пространства X и классами y , к которым они относятся. Обозначим T (множество обучающих выборок) множество непустых конечных подмножеств K , C (множество контрольных выборок) – множество непустых конечных подмножеств объектов пространства X . Будем считать, что модель зависимости m такова, что по ней однозначно может быть построено зависящее от значений параметров q детерминированное решающее правило $R_{mq} : X \rightarrow y$. Для R_{mq} обозначим $P(R_{mq})$ – вероятность правильной и точной классификации объектов из X .

Алгоритмом классификации (АК) для модели m назовем отображение $L_m : T \rightarrow Q$, которое по обучающей выборке $t \in T$ вычисляет значения параметров q , и как следствие, решающее правило R_{mq} . Будем считать, что предметная область (ПО), в которой решается задача классификации, характеризуется значениями параметров q^* . Обозначим $P_{L_m}(\mu)$ – вероятность того, что $L_m(t) = q^*$ (для дискретных значений рассмотрим вероятность совпадения значений параметров, а для числовых вещественных значений – вероятность того, что они отличаются не более чем на заданное число) для обучающих выборок t объемом μ .

Постановка задачи классификации: для модели m построить такой АК L_m , что $P_{L_m}(\mu)$ стремится к 1 при увеличении объемов μ обучающих выборок. Обозначим: $R_m(\mu)$ – множество решающих правил $R_{mL_m(t)}$ для обучающих выборок t объемом μ ; $P_{R_m}(\mu)$ – вероятность правильной и точной классификации объектов из X решающими правилами из $R_m(\mu)$. Назовем функцию $P_{R_m}(\mu)$ – внешним свойством АК L_m , а функцию $P_{L_m}(\mu)$ – его внутренним свойством.

Будем называть модель m хорошо интерпретируемой, если она является минимально интерпретируемой и, кроме того, названия и значения параметров q , а также соотношения модели m являются хорошо интерпретируемыми, т.е. понятными специалистам соответствующей ПО. Вектор значений параметров q назовем базой знаний, а АК L_m – алгоритмом индуктивного формирования знаний.

Оценки функций $P_{L_m}(\mu)$ и $P_{R_m}(\mu)$ могут быть получены в сериях компьютерных экспериментов на модельных данных, сформированных следующим образом (Рис.1). Будем считать, что для модели зависимости m может быть по-

строена зависящая от значений параметров q модель генератора G_{mq} модельных обучающих выборок. Предполагается, что G_{mq}^* является адекватной моделью генератора выборок для этой ПО. В компьютерном эксперименте по модели m случайным образом (в рамках заданных ограничений) генерируются модельные значения параметров q^* (для хорошо интерпретируемых моделей зависимости будем также называть их модельными базами знаний (МБЗ)); с помощью модели генератора G_{mq}^* генерируется подходящее множество модельных выборок различных объемов μ ; по модельным обучающим выборкам t с помощью АК L_m формируются значения параметров q' (для адекватных и хорошо интерпретируемых моделей зависимости будем также называть их индуктивно формируемыми базами знаний (ИФБЗ)), для которых вычисляются оценки функций вероятности: $P_{L_m}(\mu)$ и $P_{R_m}(\mu)$, используя множество модельных контрольных выборок c .



Рис.1. Схема организации серии компьютерных экспериментов на модельных данных

Степенью разделимости классов ПО будем называть $P_{R_{mq}}^*$ – вероятность правильной и точной классификации объектов ПО решающим правилом R_{mq}^* . Внешние свойства АК L_m будем называть абсолютно хорошими, если с ростом объема μ обучающих выборок t имеет место сходимость $P_{R_m}(\mu)$ к 1 (абсолютно плохими, если такая сходимость отсутствует). Внешние свойства АК L_m будем называть относительно хорошими, если с ростом объема μ обучающих выборок t имеет место сходимость $P_{R_m}(\mu)$ к $P_{R_{mq}}^*$ (относительно плохими, если такая сходимость отсутствует). Внутренние свойства АК L_m будем называть хорошими, если с ростом объема μ обучающих выборок t имеет место сходимость $P_{L_m}(\mu)$ к 1 (плохими, если такая сходимость отсутствует). Хорошим АК будем называть тот, который имеет хорошие внутренние свойства. Плохим АК будем называть тот, который имеет плохие внутренние свойства.

Выдвинем гипотезу о том, что внешние свойства АК в большей мере отражают степень делимости классов ПО, а его внутренние свойства, напротив, отражают именно свойства АК. Для подтверждения этой гипотезы необходимо показать, что для одной и той же модели зависимости с параметрами существуют такие хорошие и плохой АК, что: при хорошей делимости классов ПО плохой АК может иметь как абсолютно хорошие, так и абсолютно (и относительно) плохие внешние свойства; при плохой делимости классов ПО оба АК имеют абсолютно плохие внешние свойства, а плохой АК и относительно плохие внешние свойства; при любой делимости классов ПО хороший АК имеет относительно хорошие внешние свойства.

В качестве модели зависимости с параметрами m в компьютерных экспериментах использована непримитивная упрощенная онтология медицинской диагностики, представленная системой логических соотношений с параметрами. Каждое заболевание обладает «клинической картиной» (КК) – набором таких признаков, значения которых зависят от заболевания и изменяются его клиническими проявлениями. Каждое «описание клинического проявления» (ОКП) заболевания по признаку представляет собой последовательность периодов динамики (ПД) этого признака, количество которых задается значением параметра «число периодов динамики» (ЧПД). Каждый ПД характеризуется возможными значениями признака в нем, а также верхней и нижней границами своей длительности.

Вектором переменных x в этой онтологии является вектор признаков, значениями переменной y (классами) – названия заболеваний, а компонентами вектора параметров q являются неинтересные параметры «возможные значения» и «клиническая картина», а также интересные параметры «нормальные значения», «число периодов динамики», «значения для периода», «верхняя граница», «нижняя граница». Значением каждого признака в истории болезни (ИБ) является функция времени с конечной областью определения (конечным множеством моментов наблюдения, своим для каждого признака и каждой ИБ) и конечной областью значений (подмножеством возможных значений признака).

Подставляя названия классов, переменных и известные значения неинтересных параметров в систему логических соотношений с параметрами, получаем, что эта система распадается на группы логических соотношений, соответствующие парам (заболевание, признак). Любые две такие группы логических соотношений не имеют общих интересных параметров. В результате исходная задача классификации сводится к множеству задач классификации, каждую из которых будем называть частной задачей классификации для пары (заболевание y_j , признак x_i). Входными данными частной задачи классификации для пары (заболевание y_j , признак x_i) является подмножество обучающей выборки t – обучающая выборка t_{xy} , каждый элемент которой представлен функцией времени. Если признак x_i , не входит в КК заболевания y_j , то для этой пары декомпозиция задачи классификации приводит к частной задаче классификации с полной информацией. Если признак x_i , входит в КК заболевания y_j , то для этой пары решается частная задача классификации с неполной информацией.

Результат решения задачи классификации для упрощенной онтологии медицинской диагностики представляет собой объединение результатов решения частных задач классификации для всех пар (заболевание y_j , признак x_i).

В компьютерных экспериментах использовались два алгоритма для решения частной задачи классификации с неполной информацией. Алгоритм случайных расстановок границ периодов динамики сводит частную задачу классификации с неполной информацией к частной задаче классификации с полной информацией путем задания случайных (в соответствии с рядом ограничений) значений ненаблюдаемых неизвестных для каждой ИБ с диагнозом заболевание y_j обучающей выборки. Алгоритм направленного поиска расстановок границ периодов динамики в себя следующие последовательные этапы: итеративный процесс формирования корректного множества альтернатив ОКП, релевантных обучающей выборке; выбор из этого множества одной альтернативы ОКП, называемой первым приближением к ОКП; уточнение значений параметров первого приближения к ОКП – формирование ОКП заболевания y_j по признаку x_i .

Для проверки выдвинутой в гипотезы были проведены три серии компьютерных экспериментов: в 1-ой серии были получены оценки свойств алгоритма случайных расстановок границ периодов динамики в условиях хорошей разделимости заболеваний (классов ПО), которая смоделирована тем, что в МБЗ у разных заболеваний совпадающими генерировались значения параметров «клиническая картина», «число периодов динамики», а также номера критических ПД для одних и тех же признаков (при этом значения параметров «значения для периода», «верхняя граница» и «нижняя граница» для соответствующих ПД у одних и тех же признаков при разных заболеваниях генерировались случайным образом (в рамках заданных ограничений)); во 2-ой серии были получены оценки свойств обоих АК также в условиях хорошей разделимости заболеваний, которая в отличие от предыдущей серии смоделирована тем, что в МБЗ для соответствующих ПД у одних и тех же признаков при разных заболеваниях случайным образом (в рамках заданных ограничений) генерировались только значения параметров «верхняя граница» и «нижняя граница», а «значения для периода» генерировались совпадающими; в 3-ей серии были получены оценки свойств обоих АК в условиях плохой разделимости заболеваний, которая в отличие от предыдущей серии смоделирована тем, что в МБЗ для соответствующих ПД у одних и тех же признаков при разных заболеваниях случайным образом (в рамках заданных ограничений) генерировались только значения параметра «верхняя граница», а «значения для периода» и «нижняя граница» генерировались совпадающими.

Для каждой серии компьютерных экспериментов были сгенерированы 5 различных МБЗ. На основе каждой МБЗ были сгенерированы: 5 наборов модельных обучающих выборок, каждый из которых включал по одной выборке объемом от 10 до 1280 (каждый следующий объем больше предыдущего в два раза) ИБ каждого заболевания; одна модельная контрольная выборка объемом 1280 ИБ каждого заболевания. Для проведения серий компьютерных экспериментов был разработан комплекс программ, соответствующий схеме организации компьютерных экспериментов на модельных данных, предложенной в работе (Рис.1). Комплекс

состоит из подсистем: генерации модельных данных, реализации алгоритма классификации, оценки и визуализации результатов.

В компьютерных экспериментах показано (Рис.2), что для алгоритма случайных расстановок границ периодов динамики с ростом объема μ обучающих выборок t сходимость $P_{Lm}(\mu)$ к 1 отсутствует для всех интересных параметров. Также в компьютерных экспериментах показано, что для алгоритма направленного поиска расстановок границ периодов динамики с ростом объема μ обучающих выборок t присутствует устойчивая сходимость $P_{Lm}(\mu)$ к 1 для всех интересных параметров. Кроме того, в проведенных сериях экспериментов показано, что при хорошей разделимости классов ПО алгоритм случайных расстановок границ периодов динамики может иметь как абсолютно (и относительно) хорошие (1-я серия экспериментов), так и абсолютно (и относительно) плохие оценки внешних свойств (2-я серия экспериментов), а при плохой разделимости классов ПО такой алгоритм имеет абсолютно и относительно плохие оценки внешних свойств (3-я серия экспериментов). Также в проведенных сериях экспериментов показано, что вне зависимости от степени разделимости классов ПО, оценки внешних свойств алгоритма направленного поиска расстановок границ периодов динамики являются относительно хорошими, а при плохой разделимости классов ПО оценки внешних свойств такого алгоритма являются абсолютно плохими.

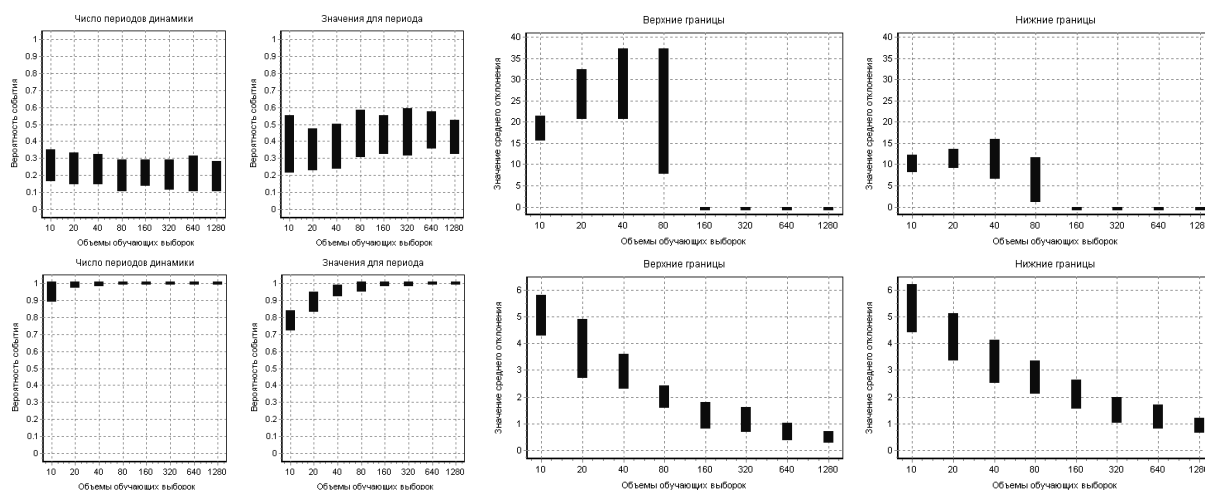


Рис.2. Оценки внутренних свойств алгоритма случайных расстановок границ периодов динамики (вверху) и алгоритма направленного поиска расстановок границ периодов динамики (внизу)

Таким образом показано, что решение задачи классификации в разработанной в диссертации постановке является и решением задачи классификации в традиционной постановке, однако обратное не всегда верно.

В третьей главе решается третья задача диссертации. В качестве адекватной и хорошо интерпретируемой модели зависимости m рассматривается онтология медицинской диагностики, приближенная к реальной, которая представлена в виде небогатой системы логических соотношений с параметрами. В этой онтологии рассматривается один вид причинно-следственных отношений – клинические проявления, каждое заболевание обладает КК. Однако, каждое ОКП заболевания по признаку, входящему в КК этого заболевания, имеет дополнительный параметр: «варианты», значением которого является множество вариантов клинического проявления (ВКП). Каждое ВКП заболевания представляет собой по-

следовательность ПД этого признака, количество которых задается значением параметра «число периодов динамики». Каждый ПД характеризуется возможными значениями признака в нем, а также верхней и нижней границами своей длительности. Вектор переменных x , значения переменной y и обучающие выборки в этой онтологии определены так же, как в упрощенной онтологии медицинской диагностики. В этой онтологии вектором параметров q является набор неинтересных параметров: «возможные значения» (для каждого признака) и «клиническая картина» (для каждого заболевания), а также интересных параметров: «нормальные значения» (для каждого признака), «варианты» (для каждого ОКП), «число периодов динамики» (для каждого ВКП), «значения для периода», «верхняя граница» и «нижняя граница» (для каждого периода динамики).

Подставляя названия классов (заболеваний y), переменных (признаков x) и известные значения неинтересных параметров онтологии в систему логических соотношений с параметрами, получаем, что эта система распадается на группы логических соотношений, соответствующие парам (заболевание y_j , признак x_i).

Любые две такие группы логических соотношений не имеют общих интересных параметров. В результате исходная задача классификации сводится к множеству частных задач, каждая из которых соответствует одной из групп соотношений. Если в паре (заболевание y_j , признак x_i) признак x_i , не входит в КК заболевания y_j , то для этой пары декомпозиция задачи классификации приводит к частной задаче классификации с полной информацией. Если в паре (заболевание y_j , признак x_i) признак x_i , входит в КК заболевания y_j , то для этой пары декомпозиция задачи классификации приводит к частной задаче, в которой неизвестно множество классов и, как следствие, неизвестно, к какому ВКП относится конкретная ИБ. Количество параметров в таком ОКП зависит от числа ВКП, а количество параметров в отдельном ВКП равно $((\text{ЧПД} - 1) * 3) + 2$, где ЧПД – это значение параметра «число периодов динамики». Из постановки этой задачи следует, что она является частной задачей кластеризации. Результат решения задачи классификации для онтологии медицинской диагностики, приближенной к реальной, представляет собой объединение результатов решения частных задач для всех пар (заболевание y_j , признак x_i), т.е. ИФБЗ.

Обозначим $m(x, q)$ модель зависимости, причем в вектор параметров q входит параметр «множество кластеров» (B), значением которого является конечное множество множеств объектов ПО. Элементы множества, являющегося значением параметра B , будем называть кластерами. Будем считать, что для каждого $b \in B$ в q входит свой набор параметров q_b (параметров кластера b), а модель зависимости m такова, что для каждого $b \in B$ может быть построен предикат Π_{q_b} , который истинен на всех объектах ПО, входящих в кластер b , и только на них.

Будем считать, что: ПО характеризуется значениями параметров q^* , в том числе значением параметра B^* , элементы $b^* \in B^*$ которого будем называть кластерами ПО; q^* таково, что кластеры B^* попарно не пересекаются.

Будем называть набор значений вектора параметров q допустимым, если множества, определяемые предикатами Π_{qb} попарно не пересекаются. Обозначим Q' – множество всех допустимых наборов значений вектора параметров q .

Алгоритмом кластеризации для модели m назовем отображение $M_m : T \rightarrow Q'$, которое по обучающей выборке $t \in T$ вычисляет допустимые значения параметров q . Для $M_m(t)$ множества, определяемые предикатами Π_{qb} , будем называть индуктивными кластерами B' . Обозначим $P_{M_m}(\mu)$ – вероятность того, что для обучающих выборок t объемом μ : между множествами индуктивных кластеров B' и кластеров ПО B^* существует взаимно-однозначное соответствие; для соответствующих индуктивного кластера $b' \in B'$ и кластера ПО $b^* \in B^*$ наборы параметров этих кластеров $q_{b'}$ и q_{b^*} совпадают, а значения одноименных параметров равны (для дискретных значений рассмотрим вероятность совпадения значений параметров, а для числовых вещественных значений – вероятность того, что они отличаются не более чем на заданное число).

Постановка задачи кластеризации: для модели m построить такой алгоритм кластеризации $M_m(t)$, что $P_{M_m}(\mu)$ стремится к 1 при увеличении объемов μ обучающих выборок. В частной задаче кластеризации для пары (заболевание y_j , признак x_i) кластерами ПО $b^* \in B^*$ являются варианты клинического проявления ОКП заболевания y_j по признаку x_i . Каждому ВКП (кластеру ПО) b^* поставлен в соответствие свой набор параметров q_{b^*} .

Будем обозначать параметры ВКП (кластера ПО) V : $ЧПД(V)$ – число ПД в ВКП V ; для номера $p \in [1, ЧПД(V)]$ будем обозначать: $ЗПД(V, p)$ – значения для ПД p в ВКП V , $ВГ(V, p)$ – верхняя граница ПД p в ВКП V , $НГ(V, p)$ – нижняя граница ПД p в ВКП V . Будем говорить, что разные ВКП (кластер ПО) V_1 и ВКП (кластер ПО) V_2 одного и того же ОКП заболевания y_j по признаку x_i различаются, если выполнено хотя бы одно из условий (в скобках указан вид различия ВКП (вид различия кластеров ПО)): $ЧПД(V_1) \neq ЧПД(V_2)$ («по ЧПД»); существует ПД с номером $p \in [1, ЧПД(V)]$, для которого имеет место $ЗПД(V_1, p) \cap ЗПД(V_2, p) = \emptyset$ («по значениям»); существует ПД с номером $p \in [1, ЧПД(V)]$, для которого имеет место $НГ(V_1, p) - ВГ(V_2, p) > 0$ («по границам»).

Входными данными частной задачи кластеризации для пары (заболевание y_j , признак x_i) является подмножество обучающей выборки t – обучающая выборка t_{xy} , каждый элемент которой представлен функцией времени. Областью определения этой функции является конечное множество моментов наблюдения признака «признак x_i » в ИБ с диагнозом «заболевание y_j », а областью значений – значения этого признака в эти моменты наблюдения.

Для частной задачи кластеризации для пары (заболевание y_j , признак x_i) будем обозначать: H_V – ИБ, соответствующую ВКП (кластеру ПО) V для пары (заболевание y_j , признак x_i); $\mathcal{CПД}(H_V)$ – число ПД признака x_i в ИБ H_V ; $\mathcal{CМН}(H_V, p)$ – число моментов наблюдения признака x_i в ПД с номером p в ИБ H_V . ИБ в обучающей выборке t_{xy} должны быть хорошо обследованными: для любого ПД с номером $p \in [1, \mathcal{CПД}(H_V)]$ должно иметь место $\mathcal{CМН}(H_V, p) \geq 2$.

Алгоритм направленного поиска расстановок границ периодов динамики для решения частной задачи кластеризации для пары (заболевание y_j , признак x_i) для онтологии медицинской диагностики, приближенной к реальной, сводится к последовательному решению следующих оптимизационных задач: обобщение всех ИБ обучающей выборки; формирование набора индуктивных кластеров; вычисление значений параметров индуктивных кластеров.

Будем называть последовательностью ПД – последовательность целочисленных интервалов временной оси, имеющих вид $(0, \beta_i)$ при $i=1$ и вид (α_i, β_i) при $i > 1$, причем для любого $i > 1$ имеет место $\beta_{i-1} = \alpha_i - 1$. Для ИБ H обучающей выборки t_{xy} для пары (заболевание y_j , признак x_i) будем обозначать: $\mathcal{CМН}'(H, p)$ – число моментов наблюдения, принадлежащих ПД с номером p ; $\mathcal{MН}'_e(H, p)$ – момент наблюдения с номером $e \in [1, \mathcal{CМН}'(H, p)]$, принадлежащий ПД с номером p ; $\text{значение}(\mathcal{MН}'_e(H, p))$ – значение признака в момент наблюдения с номером $e \in [1, \mathcal{CМН}'(H, p)]$, принадлежащий ПД с номером p .

Обозначим *возможные значения(признак)* – область возможных значений признака, для которого решается задача кластеризации. Обозначим $\mathcal{O}(H)$ – разбиение временной оси на ПД (количество которых обозначим $k(\mathcal{O}(H))$) для признака x_i и ИБ H обучающей выборки t_{xy} , удовлетворяющее следующим условиям: $k(\mathcal{O}(H)) \leq 5$; если $k(\mathcal{O}(H)) > 1$, то $\{\text{значение}(\mathcal{MН}'_i(H, p)) \mid i \in [1, \mathcal{CМН}'(H, p)]\} \cap \{\text{значение}(\mathcal{MН}'_j(H, p+1)) \mid j \in [1, \mathcal{CМН}'(H, p+1)]\} = \emptyset$ для $p \in [1, k(\mathcal{O}(H)) - 1]$; если $k(\mathcal{O}(H)) = 1$, то $\{\text{значение}(\mathcal{MН}'_i(H, 1)) \mid i \in [1, \mathcal{CМН}'(H, 1)]\} \neq \text{значения(признак)}$; для любого ПД с номером $p \in [1, k(\mathcal{O}(H))]$ имеет место $\mathcal{CМН}'(H, p) \geq 2$.

Обозначим A – набор всех разбиений $\mathcal{O}(H)$ для некоторой ИБ H для признака x_i из пары (заболевание y_j , признак x_i). Каждому l -ому разбиению $\mathcal{O}_l(H)$ поставим в соответствие значение максимальной разности между соседними моментами наблюдения, которые отнесены к смежным ПД: $r_l = \max_p (\mathcal{MН}'_1(H, p+1) - \mathcal{MН}'_{\mathcal{CМН}'(H, p)}(H, p))$, для $p \in [1, k(\mathcal{O}(H)) - 1]$.

Наилучшим будем называть такое разбиение $\mathcal{O}^*(H)$, для которого $r^* = \min_A (r_l)$. Разбиению $\mathcal{O}^*(H)$ поставим в соответствие совокупность параметров:

$\mathcal{CПД}'(H)$ – число ПД; для любого ПД с номером $p \in [1, \mathcal{CПД}'(H)]$: $\mathcal{ЗДП}'(H, p)$ – значения признака в моменты наблюдения в ПД с номером p , $\mathcal{ВГ}'(H, p)$ – верхняя граница ПД с номером p , $\mathcal{НГ}'(H, p)$ – нижняя граница ПД с номером p .

Обобщением $J(H)$ ИБ H будем называть совокупность параметров, имеющих следующие значения: $\mathcal{CПД}'(H) = k(O^*(H))$; для каждого ПД с номером $p \in [1, k(O^*(H))]$ для $O^*(H)$: $\mathcal{ЗДП}'(H, p) = \{\text{значение}(MН'_i(H, p)) \mid i \in [1, \mathcal{СМН}'(H, p)]\}$;

если $p = 1$, то $\mathcal{ВГ}'(H, 1) = MН'_1(H, 2)$, $\mathcal{НГ}'(H, 1) = MН'_{\mathcal{СМН}'(H, 1)}(H, 1)$; если

$1 < p < \mathcal{CПД}'(H)$, то $\mathcal{ВГ}'(H, p) = MН'_1(H, p+1) - MН'_{\mathcal{СМН}'(H, p-1)}(H, p-1) + 1$,

$\mathcal{НГ}'(H, p) = MН'_{\mathcal{СМН}'(H, p)}(H, p) - MН'_1(H, p)$.

Задача обобщения всех ИБ обучающей выборки состоит в том, чтобы для каждой ИБ H обучающей выборки t_{xy} сформировать ее обобщение.

Определим отношение эквивалентности между ИБ обучающей выборки t_{xy} по признаку, для которого решается задача кластеризации, следующим образом. Будем говорить, что две ИБ H^i и ИБ H^j обучающей выборки t_{xy} эквивалентны, если для их обобщений $J(H^i)$ и $J(H^j)$ имеет место $\mathcal{CПД}'(H^i) = \mathcal{CПД}'(H^j)$ и для всех ПД с номерами $p \in [1, \mathcal{CПД}'(H^i)]$ выполнены следующие условия: $\mathcal{ЗДП}'(H^i, p) \cap \mathcal{ЗДП}'(H^j, p) \neq \emptyset$, $\mathcal{ВГ}'(H^i, p) \geq \mathcal{НГ}'(H^j, p)$, $\mathcal{НГ}'(H^i, p) \leq \mathcal{ВГ}'(H^j, p)$. Будем также считать, что для данного отношения выполнено условие транзитивности, т.е. если H^i эквивалентно H^k и H^k эквивалентно H^j , то H^i эквивалентно H^j . Из определений индуктивных кластеров \mathcal{B}' и отношения эквивалентности между ИБ следует, что две ИБ, принадлежащие одному и тому же индуктивному кластеру, принадлежат одному и тому же ВКП (кластеру ПО). Эти ИБ принадлежат обучающей выборке t_{xy} , поэтому каждая из них принадлежит одному из ВКП (кластеров ПО). Из способа формирования обобщений ИБ и определения эквивалентности ИБ следует, что любому индуктивному кластеру принадлежат ИБ, соответствующие одному и тому же ВКП (кластеру ПО).

Задача формирования набора индуктивных кластеров состоит в том, чтобы разбить обучающую выборку t_{xy} на классы эквивалентности \mathcal{B}' .

Задача вычисления значений параметров индуктивных кластеров состоит в том, чтобы найти оптимальные значения всех параметров q_b , всех индуктивных кластеров \mathcal{B}' , одноименных параметрам ВКП (кластера ПО) q_{b^*} .

Свойства алгоритма направленного поиска расстановок границ периодов динамики для решения частной задачи кластеризации для пары (заболевание y_j , признак x_j) для онтологии медицинской диагностики, приближенной к реальной, были исследованы методом получения оценок внешних и внутренних свойств ал-

горитмов классификации. Для этого были проведены две серии компьютерных экспериментов: в 1-ой серии были получены оценки свойств алгоритма в условиях хорошей разделимости заболеваний (классов ПО), которая смоделирована тем, что в МБЗ разные заболевания генерировались независимо друг от друга (случайным образом в рамках заданных ограничений); во 2-ой серии были получены оценки свойств алгоритма в условиях плохой разделимости заболеваний, которая в отличие от предыдущей серии смоделирована тем, что в МБЗ у разных заболеваний (случайным образом в рамках заданных ограничений) генерировались: совпадающими значения параметра «клиническая картина» и количества ВКП (кластеров ПО) для одних и тех же признаков, у соответствующих ВКП совпадающими значения параметров «число периодов динамики» и «номер критического периода динамики», в соответствующих ПД соответствующих ВКП совпадающими значения параметров «значения для периода» и «нижняя граница», а разница между значениями параметров «верхняя граница» в диапазоне от -1 до 1 .

Для каждой серии компьютерных экспериментов были сгенерированы 5 различных МБЗ. На основе каждой МБЗ были сгенерированы: 5 наборов модельных обучающих выборок, каждый из которых включал по одной выборке объемом от 10 до 2560 (каждый следующий объем больше предыдущего в два раза) ИБ каждого заболевания; одна модельная контрольная выборка объемом 2560 ИБ каждого заболевания. Для проведения серий компьютерных экспериментов был разработан комплекс программ, соответствующий схеме организации компьютерных экспериментов на модельных данных, предложенной в работе (Рис.1).

В компьютерных экспериментах показано, что алгоритм направленного поиска расстановок границ периодов динамики для модели зависимости с параметрами (онтологии медицинской диагностики, приближенной к реальной) имеет хорошие оценки внутренних свойств во всех сериях компьютерных экспериментов, т.е. что с ростом объема μ обучающих выборок t присутствует устойчивая сходимость $P_{M_m}(\mu)$ к 1 (Рис.3). Также показано, что вне зависимости от степени разделимости классов ПО, оценки внешних свойств алгоритма классификации для модели зависимости с параметрами (онтологии медицинской диагностики, приближенной к реальной) являются относительно хорошими (Рис.4).

В четвертой главе решается четвертая задача диссертации. Разработан комплекс программ и технология его применения для индуктивного формирования баз медицинских знаний (в форме, принятой в медицинской литературе), включающий в себя следующие подсистемы: ввода ИБ (сторонняя); преобразования обучающей выборки во внутренний формат алгоритма классификации; индуктивного формирования баз знаний, реализующую алгоритм классификации (разработанный в диссертации); представления баз знаний в форме, принятой в медицинской литературе.

Обучающая выборка реальных данных, которая была использована при апробации комплекса программ, содержит 69 ИБ заболеваний раздела медицины «острый живот»: аппендицит (22, все с операциями), холецистит (17, из них 8 с операциями), панкреатит (16, все без операций), перфоративная язва желудка (14, все с операциями). Комплекс программ формирует описание заболеваний, входящих в базу знаний, состоящее из описаний клинических проявлений признаков,

входящих в КК этих заболеваний. Для каждого признака указывается количество ИБ, в которых он наблюдался, число вариантов его клинического проявления, а также описание этих вариантов. Описание варианта содержит информацию о числе ПД в нем, о значениях признака в этих ПД и о границах длительности этих ПД. При этом для каждого варианта указывается количество ИБ, которое его поддерживает, а также приводятся номера этих ИБ. Описание заболевания содержит дополнительную информацию об ИБ обучающей выборки с этим диагнозом: количество часов, прошедших с момента начала заболевания до поступления больного в клинику, была ли проведена операция (и, если да, то через сколько часов после поступления), а также количество дней, проведенных в клинике.

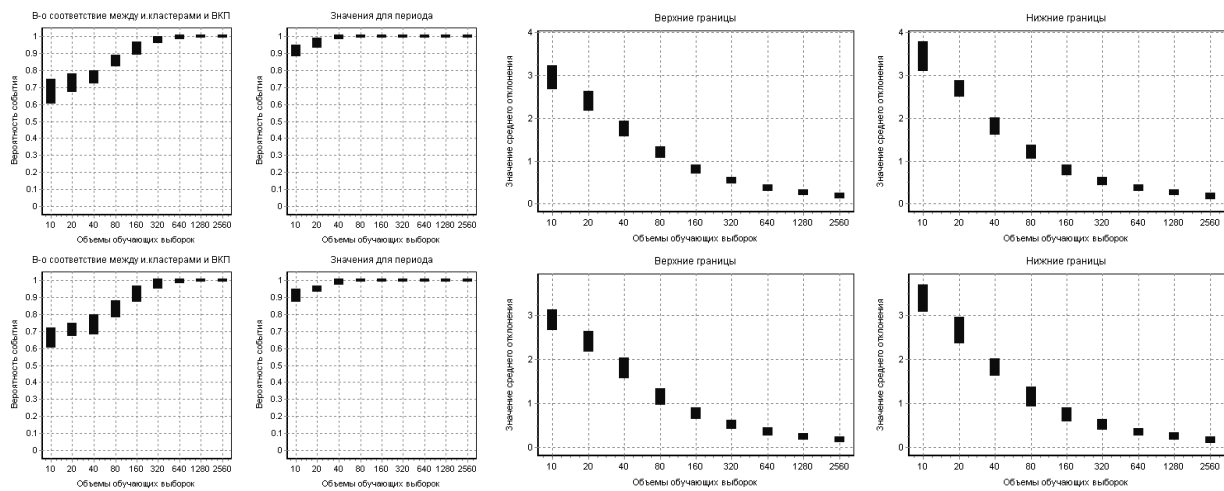


Рис.3. Оценки внутренних свойств алгоритма классификации для онтологии медицинской диагностики, приближенной к реальной, для случаев хорошей (вверху) и плохой (внизу) разделимости заболеваний

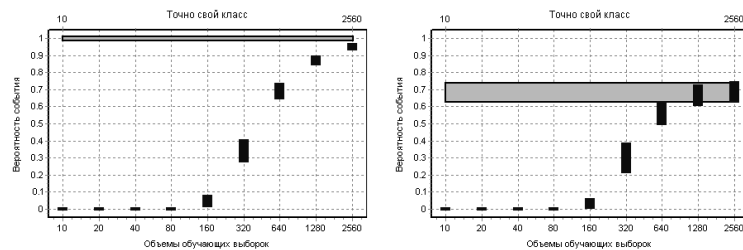


Рис.4. Оценки внешних свойств алгоритма классификации для онтологии медицинской диагностики, приближенной к реальной, для случаев хорошей (слева) и плохой (справа) разделимости заболеваний

Приведем пример описания признака «Боли в животе (Локализация)» при заболевании «аппендицит»: Количество ИБ, в которых наблюдался признак, равно 22. Число вариантов динамики равно 9. В 1-ом варианте ЧПД=2: *эпигастральная область*, затем через 5-7 часов *правая подвздошная область* (вариант поддерживают 6 ИБ). Во 2-ом варианте ЧПД=1: *правая подвздошная область* (вариант поддерживают 3 ИБ). В 3-ем варианте ЧПД=2: *весь живот, правая половина живота*, затем через 3-12 часов *правая подвздошная область* (вариант поддерживают 6 ИБ). В 4-ом варианте ЧПД=2: *нижний отдел живота*, затем через 8 часов *правая подвздошная область* (вариант поддерживают 2 ИБ). В 5-ом варианте ЧПД=3: *околопупочная область, правая подвздошная область*, через 12 часов *нижний отдел живота*, затем через 3 часа *околопупочная область* (вариант поддерживает 1 ИБ). В 6-ом варианте ЧПД=3: *эпигастральная область*, затем через 5 часов *правая подвздошная область*, затем через 4 часа *эпигастральная область* (вариант поддерживает 1 ИБ). В 7-ом варианте ЧПД=2: *весь живот, правая под-*

вздошная область, затем через 13 часов *околопупочная область* (вариант поддерживает 1 ИБ). В 8-ом варианте ЧПД=2: *правое подреберье*, затем через 9 часов *правая подвздошная область* (вариант поддерживает 1 ИБ). В 9-ом варианте ЧПД=2: *весь живот*, затем через 18 часов *правая половина живота* (вариант поддерживает 1 ИБ).

По мнению эксперта, база знаний, полученная в работе, представлена в форме, понятной практикующему врачу. При этом описания заболеваний соответствуют знаниям, имеющимся в научной и учебной медицинской литературе, а в ряде случаев дополняют их описанием динамики клинических проявлений. Формальное представление базы знаний позволяет использовать ее в экспертных системах медицинской диагностики. При этом, по мнению эксперта, ввиду высокой степени интерпретируемости полученных в работе описаний заболеваний, база знаний может быть использована при обучении студентов медицинских ВУЗов, а в случае обработки обучающей выборки большого (десятки тысяч ИБ) объема, состоящей из хорошо обследованных ИБ, сформированная на ее основе база знаний может быть использована и врачами в их профессиональной деятельности.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

1. Разработана новая постановка задачи классификации для математических моделей зависимости с параметрами (как частный случай задачи оценки значений параметров модели), в которой качество алгоритма классификации оценивается на основе близости вычисленных значений параметров модели к значениям, характеризующим предметную область.
2. Предложен метод получения оценок внешних и внутренних свойств алгоритмов классификации для моделей зависимости с параметрами с помощью компьютерных экспериментов на модельных данных. Используя этот метод, в сериях компьютерных экспериментов показано, что оценки внешних свойств алгоритмов классификации (которые связаны с постановкой задачи классификации, рассматриваемой в литературе) в большей мере отражают степень разделимость классов предметной области и в меньшей – свойства алгоритма классификации, а оценки внутренних свойств (которые связаны с постановкой задачи классификации, разработанной в диссертации), напротив, отражают именно свойства алгоритма классификации, и в меньшей мере связаны с разделимостью классов предметной области, т.е. решение задачи классификации в постановке, разработанной в диссертации, является и решением задачи классификации в традиционной постановке, однако обратное не всегда верно.
3. Разработана новая постановка задачи кластеризации для математических моделей зависимости с параметрами (как частный случай задачи оценки значений параметров модели), в которой качество алгоритма кластеризации зависит от выбранной модели зависимости с параметрами. Разработан алгоритм классификации для практически полезной и хорошо интерпретируемой математической модели зависимости с параметрами, являющейся онтологией медицинской диагностики, приближенной к реальной (заданной системой логических соотношений с параметрами), решающий задачу классификации, а также частную задачу кластеризации в новых постановках. С помощью компьютерных экспериментов на модельных данных показано, что разработан-

ный алгоритм классификации имеет хорошие оценки внутренних свойств и что вне зависимости от степени делимости классов предметной области, оценки внешних свойств этого алгоритма являются относительно хорошими.

4. Разработан комплекс программ и технология его применения для индуктивного формирования баз медицинских знаний в форме, принятой в медицинской литературе. При помощи этого комплекса, на основе обучающей выборки реальных данных, содержащей 69 историй болезни из раздела медицины «острый живот», индуктивно сформирована база знаний, представленная в форме, понятной практикующему врачу. Полученные описания заболеваний, входящие в базу знаний, по оценке эксперта, соответствуют знаниям, имеющимся в научной и учебной медицинской литературе, а в ряде случаев дополняют их описанием динамики клинических проявлений; кроме того, полученная база знаний подтверждена реальными историями болезни из обучающей выборки. Формальное представление базы знаний позволяет использовать ее в экспертных системах медицинской диагностики.

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Клещев А.С., Смагин С.В. Распараллеливание вычислений при решении задачи индуктивного формирования баз знаний. // Искусственный интеллект. – 2006. – №3. – С. 421-428.
2. Клещев А.С., Смагин С.В. Организация компьютерных экспериментов по индуктивному формированию знаний. // Научно-техническая информация. Серия 2. – М.: ВИНТИ РАН. – 2008. – №1. – С. 16-24.
3. A.S. Kleshchev and S.V. Smagin. Organization of computer experiments on inductive knowledge discovery. // Automatic Documentation and Mathematical Linguistics. – 2008. – Volume 42. – Number 1. – pp. 17-26.
4. Клещев А.С., Смагин С.В. Общий подход к проведению компьютерных экспериментов по индуктивному формированию знаний. // Программные продукты и системы. – М.: Изд-во Международный научно-исследовательский институт проблем управления. – №1. – 2008. – С. 56-58.
5. Клещев А.С., Смагин С.В. Некоторые свойства метода случайной расстановки границ периодов динамики. // Информатика и системы управления. – 2009. – №1(19). – С. 103-115.
6. Клещев А.С., Смагин С.В. Экспериментальное исследование свойств метода Монте-Карло для индуктивного формирования знаний в терминах упрощенной онтологии медицинской диагностики. // Научно-техническая информация. Серия 2. – М.: ВИНТИ РАН. – 2009. – №7. – С. 12-23.
7. A.S. Kleshchev and S.V. Smagin. Experimental study into the properties of the Monte-Carlo method for inductive formation of knowledge in terms of a simplified ontology for medical diagnostics. // Automatic Documentation and Mathematical Linguistics. – 2009. – Volume 43. – Number 4. – pp. 207-220.
8. Клещев А.С., Смагин С.В. Экспериментальное исследование свойств специализированных методов индуктивного формирования знаний для онтологий медицинской диагностики. // Материалы Всероссийской конференции с международным участием «Знания – Онтологии – Теории» (ЗОНТ-2009). – Новосибирск: ЗАО «РИЦ Прайс-Курьер». – 2009. – Том 1. – С. 137-146.

9. Клещев А.С., Смагин С.В. О роли внешних и внутренних оценок свойств методов индуктивного формирования знаний. // Научно-техническая информация. Серия 2. – М.: ВИНТИ РАН. – 2011. – №4. – С. 22-35.
10. A.S. Kleshchev and S.V. Smagin. The role of internal and external evaluation of properties of methods for the inductive formation of knowledge. // Automatic Documentation and Mathematical Linguistics. – 2011. – Volume 45. – Number 2. – pp. 91-106.
11. Клещев А.С., Смагин С.В. Выбор метода индуктивного формирования знаний для практического применения. // Материалы Всероссийской конференции с международным участием «Знания – Онтологии – Теории» (ЗОНТ-2011). – Новосибирск: ЗАО «РИЦ Прайс-Курьер». – 2011. – Том 2. – С. 15-24.
12. Клещев А.С., Смагин С.В. Задачи индуктивного формирования знаний для онтологии медицинской диагностики. // Научно-техническая информация. Серия 2. – М.: ВИНТИ РАН. – 2012. – №1. – С. 9-21.
13. A.S. Kleshchev and S.V. Smagin. Problems of inductive formation of knowledge in the ontology of medical diagnosis. // Automatic Documentation and Mathematical Linguistics. – 2012. – Volume 46. – Number 1. – pp. 8-21.

Личный вклад автора. Все результаты, составляющие основное содержание диссертации, получены автором самостоятельно. В работе [1] автором разработан алгоритм распараллеливания вычислений для задачи индуктивного формирования знаний. В работах [2-4] автором описана схема организации серии компьютерных экспериментов на модельных данных. В работах [5-11] автором проведены экспериментальные исследования алгоритмов классификации. В работах [12,13] автором предложены постановки задач классификации и кластеризации для математических моделей зависимости с параметрами, алгоритм классификации для онтологии медицинской диагностики, приближенной к реальной, а также проведено его экспериментальное исследование.

Смагин Сергей Владимирович

Метод оценки значений параметров
математической модели медицинской диагностики

Автореферат

Подписано к печати 16.07.2012
Формат 60×84/16

Усл. печ.л. 1.0
Тираж 100

Уч.–изд.л. 0.8
Заказ 25

Издано ИАПУ ДВО РАН, Владивосток, Радио, 5
Отпечатано участком оперативной печати ИАПУ ДВО РАН
Владивосток, Радио, 5