

Министерство образования Республики Беларусь

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Кафедра интеллектуальных информационных технологий

М. Д. СТЕПАНОВА, С. А. САМОДУМКИН

**ПРИКЛАДНЫЕ ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ
В ОБЛАСТИ МЕДИЦИНЫ**

Учебно-методическое пособие
для студентов специальности "Искусственный интеллект"

Минск 2000

УДК 007:681.51

ББК 32.813

С 79

Авторы:

М. Д. Степанова, С. А. Самодумкин

Степанова М.Д., Самодумкин С.А.

Прикладные интеллектуальные системы в области медицины: Учебно-методическое пособие.– Мн.: БГУИР, 2000. - с. :ил. 5

ISBN 985-444-046-X

В учебно-методическом пособии описываются основные задачи обработки информации и реализующие их компьютерные медицинские системы, анализируются их функции и структура, рассматриваются задачи интеллектуализации медицинских систем, приводится пример описания предметной области, рассматриваются математические методы классификации, используемые для постановки диагноза в медицинских интеллектуальных системах.

Пособие предназначено для курса "Прикладные интеллектуальные системы в области медицины".

УДК 007:681.51

ББК 32.813

ISBN 985-444-046-X

© М. Д. Степанова, С. А. Самодумкин 2000

СОДЕРЖАНИЕ

1. ОБЩАЯ ХАРАКТЕРИСТИКА МЕДИЦИНСКИХ СИСТЕМ.....	1
1.1. Классификация современных медицинских систем.....	4
1.2. Структура и основные компоненты медицинских систем	12
1.3. Интегрированная среда функционирования современных медицинских систем.....	24
1.4. Типология задач интеллектуализации медицинских систем	25
1.5. Формализация описания предметной области для медицинских интеллектуальных систем	27
2. МАТЕМАТИЧЕСКИЕ МЕТОДЫ ПОСТАНОВКИ ДИАГНОЗА, ИСПОЛЬЗУЕМЫЕ В МЕДИЦИНСКИХ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ	33
2.1. Математические методы решения задач медицинской диагностики	34
2.2. Классификация при наличии обучающих выборок.....	35
ЛИТЕРАТУРА.....	38

1. ОБЩАЯ ХАРАКТЕРИСТИКА МЕДИЦИНСКИХ СИСТЕМ

1.1. Классификация современных медицинских систем

Медицинская компьютерная система – это любая система, которая обеспечивает компьютерную реализацию задач медицинского назначения. Такая система, как правило, представляется совокупностью программ, предназначенных для решения задач определенного типа и объединяемых общим интерфейсом. Здесь под типами задач подразумеваются как прикладные задачи, так и типы обработки информации, основывающиеся на определенных структурах и способах хранения и представления данных.

К числу медицинских систем (МС) могут быть отнесены медицинские информационные системы, системы обработки данных (СОД), системы управления в здравоохранении; диагностические системы, информационно-справочные системы и т.п.

Рассмотрим одну из распространенных классификационных схем МС в соответствии с их назначением. Дадим характеристику МС с точки зрения реализуемых ими задач. Несмотря на большие различия медицинских систем, можно выделить ряд базовых задач. Определим сначала эти задачи, а затем опишем каждый из типов МС набором наиболее характерных для них задач.

Основные медицинские задачи. Обычно в одной МС можно выявить несколько решаемых задач. Однако в большинстве случаев можно установить одну основную (конечную) задачу.

Статистическая обработка данных используется почти в каждой МС, однако, она чаще всего не является конечной целью, а лишь служит одним из этапов обработки информации. Статистическая обработка – конечная цель для систем обработки данных, используемых в научных исследованиях, массовых профилактических обследованиях, для первичной обработки диагностической информации и т. п.

Информация медицинского персонала как конечная цель характерна для медицинских информационных систем. Под информационной целью понимается сбор, первичная обработка, сжатие информации, ее упорядочение, хранение и выдача потребителю в нужном виде. Медицинские информационно-справочные системы могут хранить компьютерные истории болезни, разнообразные сведения

библиографического характера, справочные материалы, результаты ведущихся и выполненных научных исследований (клинических или экспериментальных), различную информацию административно-управленческого характера, материалы для годовых, месячных и других отчетов, сводок и т. д. Например, система CCIS (COMPUTERIZED CLINICAL INFORMATION SYSTEM) [1] содержит программное обеспечение и базы данных для несчастных случаев, в том числе отравлений наркотиками, ядами, лекарственными препаратами (идентификация, оценка доз, диагностика, лечение).

Медицинское распознавание, т. е. отнесение патологических состояний, процессов или биологических объектов к одному из классов, является наиболее часто встречающейся задачей в разрабатываемых и эксплуатируемых медицинских системах.

Вариантами медицинского распознавания являются клиническая диагностика, прогнозирование течения процесса, идентификация каких-либо веществ или микроорганизмов, скрининг, т. е. выделение больных или лиц с повышенным риском заболевания при массовых профилактических осмотрах, отбор или выделение какой-либо части популяции в соответствии с заданными критериями.

Частными задачами, которые могут быть отнесены к этапам решения задач медицинского распознавания, являются: оценка информативности признаков и отбор наиболее информативных из них, разработка или выбор алгоритмов распознавания, создание "диагностических нормативов" в виде таблиц, содержащих данные о вероятностях симптомов и априорных вероятностях болезней.

Контроль состояния организма является вариантом медицинского распознавания. Речь идет о текущей оценке состояния, например, в послеоперационном периоде, при лечении ишемической болезни сердца, при тяжелых травмах и угрожающих состояниях другого происхождения, при беременности и родах. Чаще всего эти модели используются в системах, предназначенных для непрерывной обработки с помощью ЭВМ электрокардиограмм, электроэнцефалограмм, кривых дыхания, пульса, артериального давления и т. п.

Наиболее частой целью обработки информации при контроле состояния пациента является быстрое распознавание небольшого числа состояний, например

нормального и ненормального с подачей сигнала тревоги, вызовом дежурного персонала, автоматическим включением регистрации данных или проведением неотложных мероприятий. Например, байесовская вероятностная модель применяется для отнесения больных, находящихся в палате интенсивной терапии после операции на открытом сердце, на основе мониторингового наблюдения за пульсом и артериальным давлением к одному из трех классов: I – относительное благополучие; II – состояние, требующее усиления наблюдения; III – опасное состояние, требующее срочного врачебного вмешательства.

Сравнение с физиологической нормой, контроль за развитием здорового организма и массовые профилактические осмотры. Сравнение с физиологической нормой необходимо при массовых профилактических осмотрах для выявления факторов риска тех или иных заболеваний, уровня заболеваемости и инвалидности, для оценки типа индивидуального развития у детей, для физиологической и гигиенической оценки коллективов. Часто в этих случаях речь идет о моделях, описывающих границу между нормой и патологией и помогающих определить вероятность перехода этой границы. Например, при массовых профилактических осмотрах выделяют группу "повышенного риска", скажем риска кардиологического заболевания. В этой группе вероятность заболевания существенно выше этой вероятности среди остальных обследованных. Люди, попавшие в такую группу, подлежат дальнейшему углубленному осмотру и наблюдению.

Обеспечение лечебного процесса. Это, по существу, большая группа довольно разнородных по своим задачам моделей, предназначенных для простого расчета лечебных дозировок и рецептов, например, при инфузионной терапии; для определения оптимальной лечебной тактики, например, на основе теории игр и статистических решений, методов оптимизации; для контроля процесса лечения – введения лекарств, работы лечебного оборудования; для автоматизации самого процесса лечения, например, применения биоуправляемых дозаторов лекарств, управления протезами органов и физиологическими функциями; поддержания сердечного ритма, регулирования массы циркулирующей крови, управления аппаратами искусственного кровообращения, дыхания, автоматического наркоза и т. д.

Администрация и управление здравоохранением (медицинским обслуживанием). В административно-управленческих моделях речь идет об

автоматизации таких процессов, как учет, отчетность, экономическое планирование, распределение ресурсов (людских, материальных), управление медицинскими запасами, аптечным снабжением, лабораторным обслуживанием, потоками документов, хранением и выдачей административной медицинской информации.

Для реализации этих задач предназначены системы планирования, контроля, оценки работы медицинских подразделений, учреждений, управления медицинскими коллективами. В последнее время такие системы получили название офисных медицинских систем. К этому классу будем относить системы, обеспечивающие информационную поддержку функционирования лечебного учреждения (ЛУ). Приведем примеры офисных медицинских систем.

Система организации приема больных MEDICAL MANAGEMENT SYSTEM SCHEDULING [1] составляет и поддерживает графики работы врачей, использование помещений и оборудования; назначает больным время приема у врача, исходя из анализа различных условий и пожеланий. Специальный модуль постоянно следит и обновляет информацию о том, сколько времени пациент ждет приема с момента своего прихода в клинику. Система ежедневно распечатывает графики (расписания), следит за несостоявшимися посещениями, регистрирует новых пациентов при первом посещении, хранит полную информацию о больном и имеет возможности поиска в ней, выдает отчеты об использовании врачом рабочего времени с учетом неявки больных.

Система распределения ресурсов THE RESOURCE SCHEDULER [1] позволяет планировать работу 200 врачей и 999 единиц ресурсов (ассистентов, оборудования и т.п.). Основной график определяет стандартные рабочие часы для каждого врача в незанятое время. Отмененные приемы фиксируются автоматически.

Система INFOSTAT [1] поддерживает обработку административной, финансовой и исполнительной информации следующего вида: истории болезни, учет больных, планы медицинского ухода за больными, лекарственные препараты, бухгалтерский учет, счета к оплате, инвентарь, учет расходов, кадры, контроль возможных выплат, проверка утилизации.

THE MEDICAL MANAGER [1] помогает организовать работу практикующих врачей (одновременно может обслуживать от 1 до 999 врачей). Возможности системы: счета, страховые запросы, распорядок приема пациентов, девять типов

историй болезни, первое и второе извещения о повторном вызове, справочная информация, больничные обходы, медицинские карты. Записи ведутся и пополняются автоматически. Возможна безбумажная отправка страховых запросов в страховые центры. Генератор отчетов может выдать врачу-пользователю практически любой отчет по данным в системе.

Q.D. System [1] выполняет следующие функции: заполнение и распечатка карты больного; составление отчетов о процедурах; выписка рецептов; проверка на аллергические реакции и совместимость медикаментов; ввод информации с голоса непосредственно для занесения в карту; дистанционный доступ к данным; ведение бухгалтерии; составление расписаний.

Таким образом, офисные медицинские системы обеспечивают: составление графиков работы врачей; планирование и учет использования помещений и оборудования; составление отчетов об использовании врачом рабочего времени; ведение историй болезни; учет больных; учет лекарственных препаратов; бухгалтерский учет; выписку счетов к оплате; учет инвентаря; учет расходов; контроль возможных выплат; проверку утилизации.

Обучение медицинского персонала и больных, первых – процессам диагностики, выполнения медицинских манипуляций, планирования и организации лечения, управлению здравоохранением, вторых – новым навыкам, необходимым для восполнения утраченных функций. Система MAXX (MAXIMUM ACCESS TO DIAGNOSIS AND THERAPY) [1] содержит данные из основных пособий по медицине (пересматриваемые каждые четыре месяца) о проверенных принципах лечения в современной медицинской практике и может использоваться в качестве источника для диагностических протоколов, справочника для проверки правильности лечения, обучающего средства для медиков.

Многоцелевые системы. Они обычно позволяют решать задачи статистической обработки, информации персонала, распознавания, планирования лечения и управления здравоохранением.

Управление системами медицинских приборов и аппаратов. Интенсивное развитие медицинского приборостроения привело к созданию компьютерных систем для лабораторных исследований. К ним относятся лабораторные системы микробиологии, радиологии, рентгенографии, компьютерной томографии, ультразвукового исследования и др.

3M LABORATORY INFORMATION SYSTEMS [1] – это используемое в различных больницах семейство современных компьютерных систем. В него входят: система 3M LIS для общелабораторных и микробиологических отделений, система 3M банка крови для служб переливания крови, система 3M анатомической патологии и цитологии.

DYNAPULSE [1] – система для измерения систолического, диастолического и среднего артериального давления и пульса. Она графически изображает форму волны артериального давления, а также историю и тенденции кровяного давления и пульса пациента; имеет базу данных, содержащую записи о каждом пациенте.

Системы для лабораторных исследований выполняют следующие функции: ввод и хранение информации о микробиологических, бактериологических, иммунологических исследованиях; выдачу результатов обследования пациентов.

К недостаткам систем для лабораторных исследований относятся: их практически полная автономность; отсутствие интерфейса с офисными системами, в частности с ведущими истории болезни; отсутствие взаимодействия с базами данных, с диагностическими системами; отсутствие средств интерпретации результатов лабораторных исследований.

Диагностика и принятие решений. С момента появления первых работ, посвященных компьютерным методам диагностики и принятия медицинских решений, предложено большое число математических моделей в разных областях медицины, причем многие из них были проверены на практике и показали достаточно высокую эффективность. Среди них следует отметить: 1) численные модели, основанные на теореме Байеса, отношении правдоподобия, методах дискриминантного, кластерного, факторного анализа, теории решений, алгоритмах модельных нейронных сетей; 2) логические модели, использующие булеву логику для выражения функциональных взаимосвязей и отношений между симптомами, проявлениями заболеваний и самими заболеваниями. Использовались также возможности выражения медицинских знаний в терминах логики предикатов, многозначной и модальной логики.

В последние 10 – 15 лет для решения задач диагностики и систем принятия решений в медицине широко привлекаются методы искусственного интеллекта. Их компьютерная реализация осуществляется в виде экспертных систем (ЭС). Медицина относится к числу тех областей деятельности, где применение ЭС может дать существенный эффект. Большие диагностические и лечебные

возможности экспертных систем связаны с тем, что их базы знаний могут включать знания большого числа специалистов о возможных диагностических признаках того или иного заболевания, видах его проявления, ходе его протекания, о взаимосвязях с факторами смежных областей; учитывать особенности конкретных пациентов, поддерживать принятие решений в критических точках лечебного процесса и т. д. и в итоге накапливать большой объем сведений о заболеваниях и их формах, превышающий знания одного человека (или одной школы). Эти свойства экспертных систем делают их эффективным диагностическим средством, гарантирующим учет всех диагностических признаков и анализ влияния всех факторов, что практически исключает ошибки при построении диагноза [2, 3].

Приведем примеры таких систем.

Система MYCIN была спроектирована для ассистирования врачам в диагностике и назначении курса лечения при инфекционных церебральных заболеваниях. Медицинские знания о заболеваниях представлялись в виде продукционных правил "ЕСЛИ – ТО" в сочетании с коэффициентами неопределенности. Всего в систему было включено около 500 правил.

В системе CASNET (диагноз и прогнозирование течения глаукомы) рассматриваются ассоциативные связи между булевыми комбинациями результатов различных исследований и единственным патофизиологическим состоянием, которое оценивается с некоторым коэффициентом доверия. В CASNET используются правила, связывающие с заданным правдоподобием эффект s_j , следующий в результате появления причины s_i . Кроме того, были предложены правила классификации, позволяющие по булевым комбинациям подтверждения или неподтверждения состояний указать диагностическое утверждение.

Система EXPERT позволяет задать продукционные правила с определенным уровнем доверия и иерархию этих правил. Эта система используется в ревматологии, эндокринологии и офтальмологии.

Система INTERNIST предназначена для диагностики внутренних заболеваний (более 500 заболеваний). В этой системе используются взаимосвязи двух видов: симптом – заболевание (частота встречаемости) и заболевание – симптом (интенсивность проявления).

MYCIN, EXPERT, CASNET являются логическими системами в том смысле, что взаимосвязи между всеми видами наблюдений за пациентами формируют основу вывода "ЕСЛИ – ТО", но сам вывод не обязательно имеет строго определенный характер, который связан с коэффициентами доверия или коэффициентами неопределенности.

Перечисленные выше системы считаются в настоящее время классическими в своей области и применяются в учебных целях.

Медицинская система ILIADF AND ILIAD ACCESS – экспертная система, предназначенная для обучения и поддержки врача. Работает в двух режимах: как эксперт-консультант, предлагающий дифференциальный диагноз, и как модель пациента для обучения и тестирования (число моделируемых случаев заболевания потенциально не ограничено). База знаний системы содержит сведения более чем о 1200 внутренних расстройствах и приблизительно 6000 признаков (проявлений) болезней, что покрывает большинство известных заболеваний в следующих областях: кардиология, пульмонология, гастроэнтерология, эндокринология, неврология, инфекционные болезни, ревматология и гематология.

CLINDERM DERMAYOLOGIC DIFFERENTIAL DIAGNOSIS – аналитическая интерактивная система дифференциальной диагностики заболеваний кожи. Диагноз ставится на основе комбинации выраженных морфологических аномалий и их расположения на теле. Программа предлагает список признаков для выявления их у пациента. Она может быть использована и для проверки предполагаемого диагноза: для любого заданного диагноза она выдаст дифференциальный диагноз в различных контекстах и описание возможных признаков заболевания.

CAPTURE – экспертная система, служащая для регистрации случаев обнаружения рака и организации диспансерного учета больных в соответствии с рекомендациями Американского общества раковых заболеваний: автоматически следит за прохождением необходимых обследований, рассылает напоминания больным, хранит результаты анализов и устанавливает сроки последующих обследований.

Система MedTrac поддерживает ведение амбулаторной карты больного, используя базу знаний экспертной системы, охватывающую более 3000 диагнозов. Система осуществляет диагностический анализ и анализ риска по картам больных. Результаты лабораторных исследований заносятся

автоматически из системы лабораторной поддержки ASTM-1238. SQL-поиск позволяет находить в базе данных группы больных с одинаковыми комбинациями симптомов и другую информацию. Для автоматического порождения текстов по картам система пользуется как медицинской терминологией, так и общеупотребительной лексикой. Кроме того, MedTrac автоматизирует выписку рецептов, памяток для пациентов, отслеживает коды диагнозов, необходимые для оплаты лечения и статистики, поддерживает интерфейс с головной базой данных.

К недостаткам большинства экспертных систем относятся их узкая специализация, отсутствие режима on-line работы с медицинским оборудованием, отсутствие интерфейса с лабораторными и офисными системами и между самими ЭС.

К числу задач, решаемых с помощью медицинских компьютерных систем, следует также отнести **планирование медицинского эксперимента и анализ медицинской литературы**. Математические модели этих задач и способы их реализации рассмотрены в работе [4].

1.2. Структура и основные компоненты медицинских систем

Структуру и основные компоненты медицинских систем определим, исходя из требований, которые необходимо учитывать при проектировании компьютерных систем лечебного учреждения. Проектирование таких систем предполагает предварительный анализ задач, решаемых в медицинской практике (см. подраздел 1.1).

Любые работы в области информатизации лечебно-диагностического процесса должны быть направлены на существенное повышение качества оказания медицинской помощи. В связи с этим применение компьютерных систем поддержки лечебно-диагностического процесса должно обеспечивать:

повышение качества, точности и производительности лабораторных исследований;

успешное и достоверное решение вопросов диагностики, профилактики и выбора тактики лечения;

увеличение эффективности использования коечного фонда ЛУ;

своевременное и качественное снабжение ЛУ лекарственными средствами, лечебным питанием и медицинским имуществом;

оперативный и всесторонний анализ деятельности ЛУ;

повышение качества подготовки медицинского персонала;
проведение научных исследований и внедрение их результатов в клиническую практику.

Для решения перечисленных выше задач с помощью медицинских компьютерных систем необходимо предусмотреть реализацию разнообразных функций ЛУ. Такие функции систематизированы в табл. 1, а соответствующая функциональная структура представлена на рис. 1. Таким образом, система, обеспечивающая решение перечисленных задач, должна проектироваться как интегрированная информационная система.

Очевидно, что решение столь разнородных задач затруднено в рамках одной системы. Для реализации указанных функций целесообразна интеграция в информационную систему лечебного учреждения следующих систем [1]:

офисной медицинской системы;
аппаратно-программных комплексов для проведения лабораторных исследований;
экспертных систем мониторинга, диагностики и прогнозирования;
системы интерпретации и обработки результатов научных исследований;
интеллектуальной обучающей системы.

Структурная схема информационной системы ЛУ представлена на рис. 2.

Рассмотрим назначение и функции каждой из системы информационной системы ЛУ.

Офисная медицинская система. Назначение этой системы – обеспечение информационной поддержки лечебного учреждения, включающей обработку административной, финансовой и исполнительной информации: ведение историй болезней, учет больных, бухгалтерский учет, учет лекарственных средств, составление отчетных документов.

Основные классы задач информационного обеспечения и функции ЛУ

Класс	Функции
1	2
<p>Информационная поддержка работы ЛУ</p>	<p>Регистрация пациентов Ведение БД по всем аспектам пребывания пациентов в ЛУ Автоматизированное ведение истории болезни Хранение и предоставление результатов функциональных, лабораторных и рентгенологических исследований Формирование и выдача медицинских заключений Составление графиков работы врачей Ведение графиков работы медицинского персонала всех уровней Планирование и учет использования помещений и оборудования Назначение больным времени приема у врача Проведение анализа работы отделения и ЛУ Ведение, обработка и анализ медицинских и хозяйственных статистических данных Учет больных Учет лекарственных препаратов Бухгалтерский учет Учет расходов Подготовка отчетных документов</p>
<p>Информационная поддержка лабораторных исследований</p>	<p>Автоматическая регистрация данных лабораторных исследований Ввод в систему и хранение данных лабораторных исследований Обеспечение удобного доступа к данным лабораторных исследований и результатам их обработки Анализ данных лабораторных исследований, обеспечение интерфейса с интеллектуальными системами для интерпретации данных и оценки тяжести состояния пациента</p>

1	2
Информационная поддержка диагностики, прогнозирования, мониторинга и лечения	Интеллектуальные системы диагностики, прогнозирования и мониторинга Интеллектуальные системы поддержки принятия решений Интерфейсы с БД и системами для лабораторных исследований
Информационная поддержка научно-исследовательских работ	Статистический анализ результатов научных исследований Интерфейсы с существующими статистическими пакетами (STATISTICA, SSPS, SYSTAT или BMDP) Интерфейсы с геоинформационными системами для обеспечения экологического мониторинга
Информационная поддержка интеллектуализации обучения	Учебные мультимедийные модули в области клинической медицины Справочная информация о заболеваниях, о методах лечения, схемы лечения и т. д.

Системы для лабораторных и диагностических исследований. К ним относятся компьютерные системы микробиологии, радиологии, диагностические системы рентгенографии, компьютерной томографии, ультразвукового исследования и другие, представляющие собой аппаратно-программный комплекс. Для обеспечения ввода информации от медицинского оборудования в компьютерную среду и последующей обработки полученных результатов, аппаратно-программный комплекс проведения лабораторных исследований должен состоять из подсистемы ввода и аналого-цифрового преобразования информации и подсистемы обработки данных.

Информация, получаемая в процессе медицинских исследований, представляется в аналоговом виде. Такие сигналы поступают в подсистему ввода и аналого-цифрового преобразования информации, где она преобразуется в последовательность цифровых кодов. Оцифрованная информация далее поступает в подсистему цифровой обработки данных, которая необходима, например, для улучшения качества изображения. Далее обработанные данные поступают в ЭВМ.



Рис. 1. Функциональная структура медицинской системы ЛУ

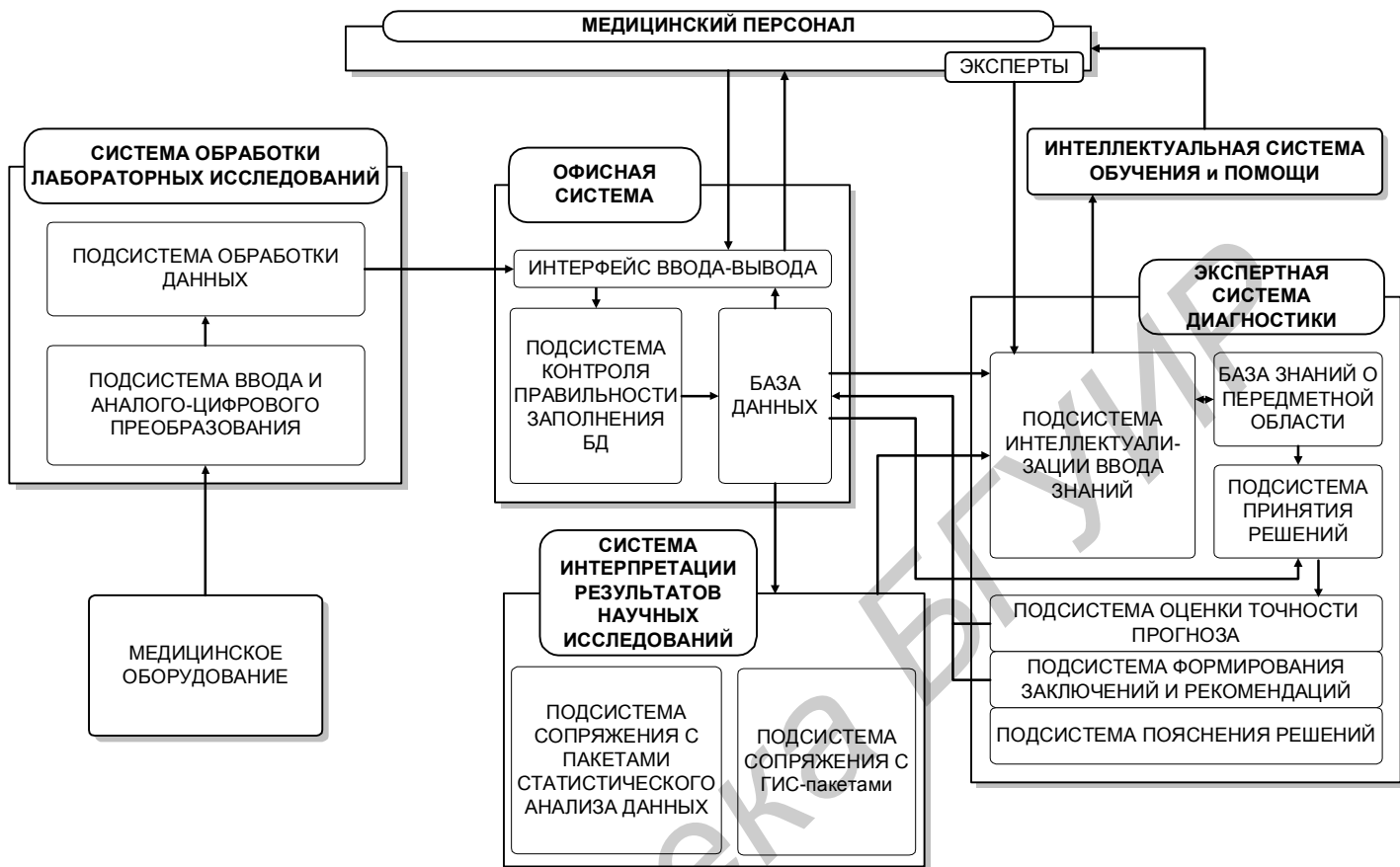


Рис. 2. Структурная схема медицинской системы ЛУ

Экспертная система для интерпретации данных исследований, диагностики, прогнозирования и мониторинга. Такая система состоит из информационных (база знаний о предметной области, база данных, хранящая сведения об изучаемых объектах) и программных компонентов: *подсистемы интеллектуализации ввода знаний, подсистемы принятия решений, подсистемы оценки точности решения, подсистемы формирования заключений и рекомендаций, подсистемы пояснения решений.*

Подсистема интеллектуализации ввода знаний (интеллектуальный конструктор) предназначена для приобретения экспертных знаний. Под приобретением знаний будем понимать, во-первых, выявление знаний из источника и, во-вторых, перенос их в компьютерную систему. В работе [5] подробно описаны способы, методы и стратегии приобретения знаний из различных источников (экспертов, группы экспертов, текстов, примеров, баз данных).

В качестве примера рассмотрим представление знаний о хронической почечной недостаточности (ХПН) и приведем описание диагноза "хронический пиелонефрит" в табличном виде (табл. 2) и представление его в виде однородной семантической сети (рис. 3, рис. 4, рис. 5) [6]. На этих рисунках показана иерархия знаний о диагнозах ХПН и их формах проявления, а также представлены экспертные знания. В частности, указывается, что множество диагнозов (узел "Диагноз") состоит из следующих диагнозов: "Хронический пиелонефрит", "Хронический гломерулонефрит", "Поликистоз почек" (рис. 4).

Определение фазы течения и функции почек показано в виде графа и/или, представленного в виде семантической сети (рис. 3). В соответствии с этим представлением в зависимости от функционального состояния почек (узел "Функция почек") и стадии лечения (узел "Фаза лечения") возможны семь различных схем протекания одного и того же заболевания. Для каждого диагноза определен симптомокомплекс, выделены количественные и качественные показатели ("Характеристика симптома"), определены их коэффициенты информативности (значимости), указано количество симптомов, для количественных показателей определены пределы измерения (верхнее и нижнее значение показателя), а для качественных показателей заданы лингвистические шкалы (рис. 4). Для каждого варианта протекания заболевания определены эталонные показатели, представляющие собой интервал варьирования в пределах измерения признака (для количественных показателей) и в диапазоне [0,100] (для качественных показателей) (рис. 5).

Первичный хронический пиелонефрит. Состояние почек - не нарушено.

Симптомы и их коэффициенты информативности	Лингвистические шкалы	Числовой диапазон	Ремиссия				Обострение					
			1	2	3	4	1	2	3	4	5	
Лабораторные показатели 0,9	Нормализация	0 – 25	•				•					
	Улучшение	25 – 50		•	•			•				
	Медленное улучшение	50 – 75				•			•	•		
	Ухудшение	75 – 100										•
Болевой синдром 0,9	Отсутствует	0 – 20	•	•								
	Слабовыраженный	20 – 50			•	•	•	•				
	Сохраняющийся	50 - 75							•	•		
	Стойкий	75 – 100										•
Артериальное давление 1,0	Нормальное	0 – 20										
	Невысокое	20 - 30	•	•								
	Снижение	30 – 40			•		•					
	Снижение без нормализации	40 – 55				•		•				
	Нестойкое	55 – 65							•			
	Повышение	65 – 85									•	
	Стабильно высокое	85 – 100										•
Анемия 1,0	Отсутствует > 120 г/л	0 – 20	•									
	Стабилизация	20 – 40		•	•		•					
	Легкая 90 – 115 г/л	40 – 60				•		•	•			
	Нарастание	60 – 80								•		
	Резко выраженная < 80 г/л	80 – 100										•
Ацидоз метаболический 0,9	Снижение	0 – 20	•									
	Стабилизация	20 – 40		•	•							
	Компенсированный	40 – 55				•	•					
	Медленно прогрессирующий	55 – 75						•	•			
	Стойкий	70 – 80								•		
	Декомпенсированный	80 – 100										•
Скорость клубочковой фильтрации 1,0	Диапазон варьирования [5,55]	5 – 7										
		8 – 12										
		12 – 15									•	•
		15 – 25										
		25 – 30							•			
		25 – 35			•	•		•				
		30 – 40		•			•					
		40 – 55	•									

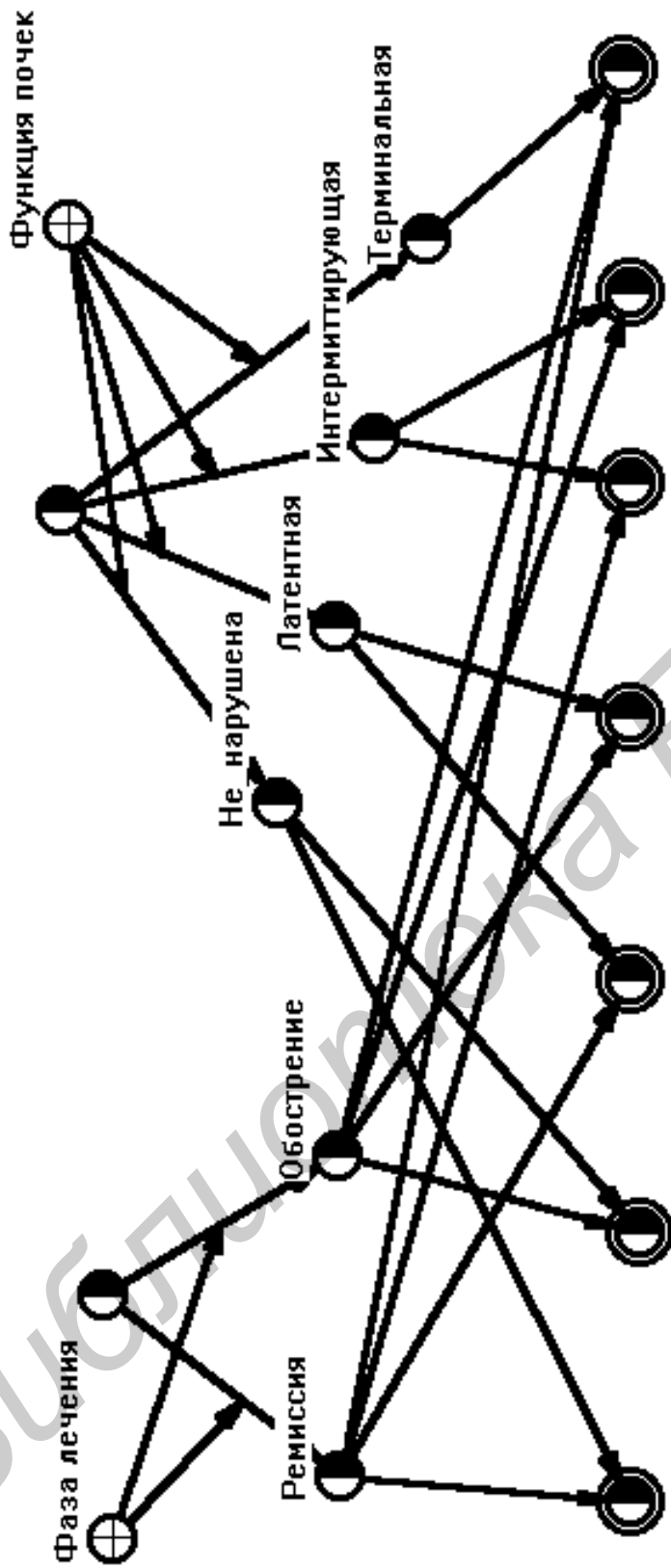


Рис. 3. Определение варианта протекания заболевания.

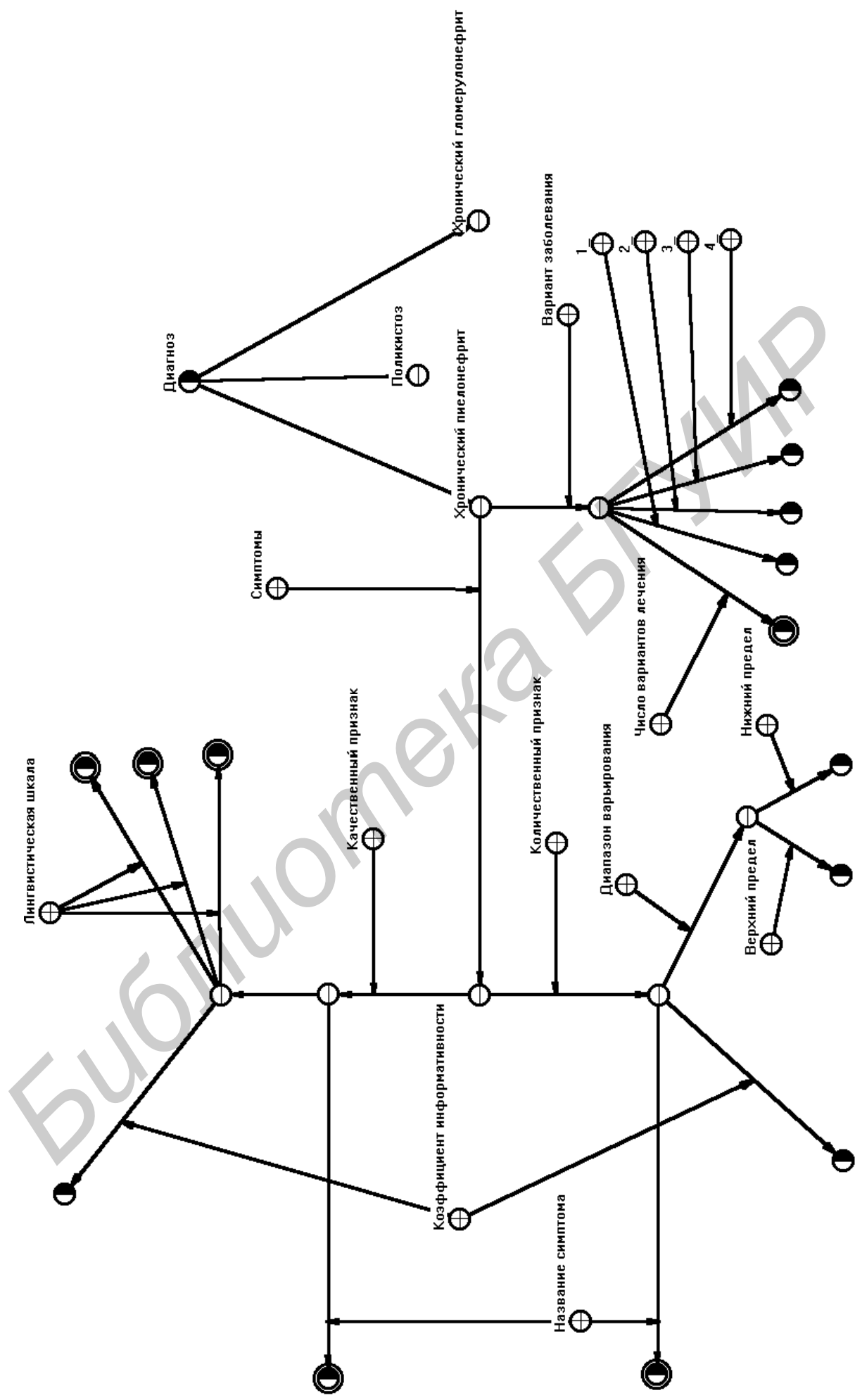


Рис. 4. Представление симптомокомплекса

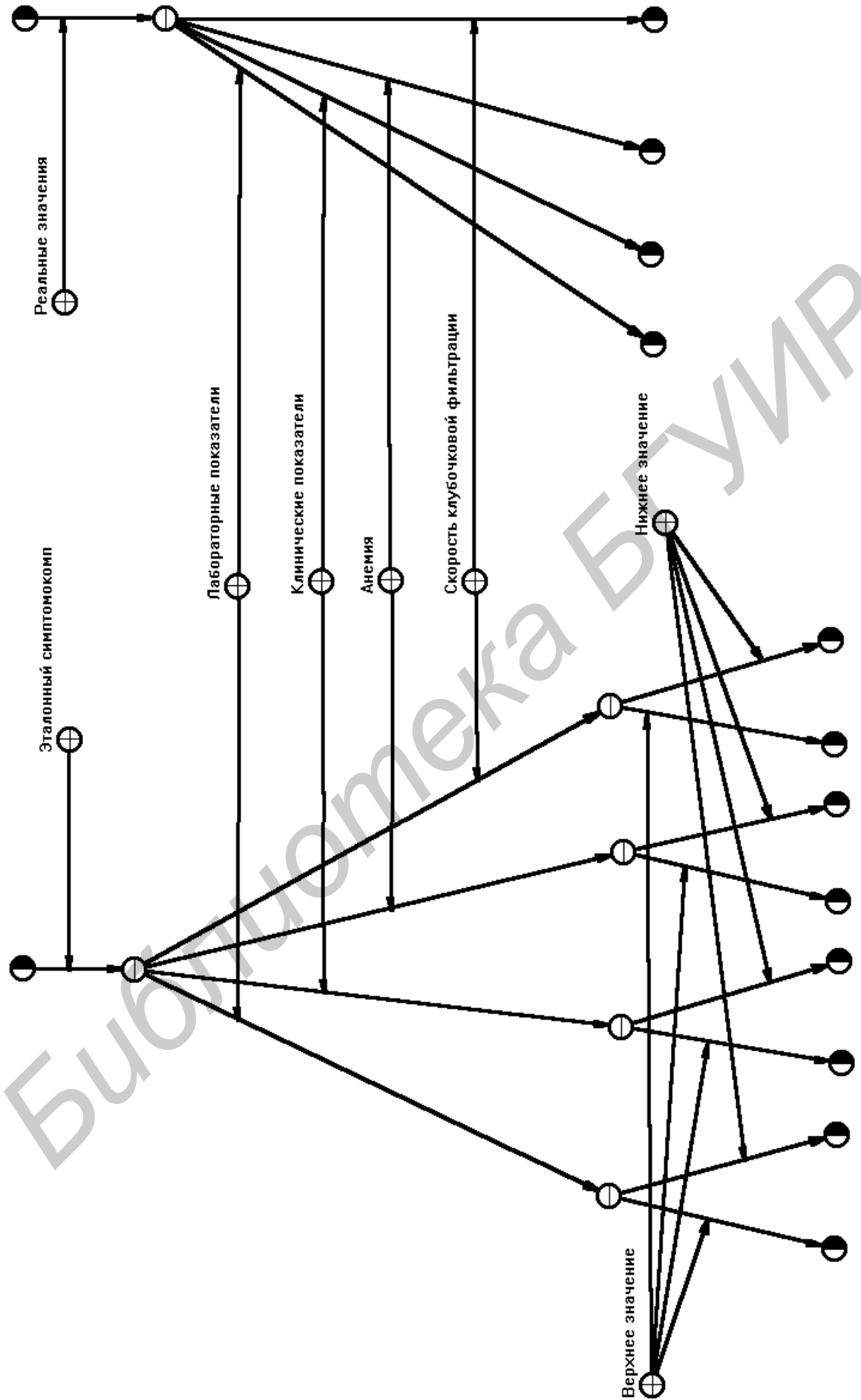


Рис. 5. Фрагмент базы знаний.

Подсистема принятия решений (решатель) производит выбор решения в конкретной ситуации. Для этого используются различные механизмы рассуждений и методы.

Подсистема оценки точности решения предназначена для оценивания достоверности полученного решения. Такая оценка позволяет лицу, принимающему решение (ЛПР), принять или отвергнуть предложение ЭС. В качестве ЛПР выступает лечащий врач. В условиях противоречивости данных или полученных выводов система должна отказаться выдать решение и сформировать соответствующее сообщение пользователю.

Подсистема формирования заключений и рекомендаций. Назначение этой подсистемы – выбор из базы знаний заключений, рекомендаций, стратегий лечения в зависимости от полученного решения.

Подсистема пояснения решений. Важнейшей особенностью экспертной системы является необходимость объяснения выбора решения и того, каким образом оно было достигнуто.

Система интерпретации и обработки результатов научных исследований. Одним из основных методов получения новых знаний, выявления взаимосвязей между различными объектами предметной области является статистическая обработка данных, накапливаемых в базе данных. Совместное применение методов статистического анализа данных, распознавания образов и экспертных знаний позволяет на более высоком и качественном уровне приобретать знания в медицине.

Ряд заболеваний обусловлен факторами внешней среды, поэтому для правильной диагностики необходимо учитывать воздействие среды обитания человека на его здоровье, например, влияние концентраций загрязняющих веществ в атмосфере или уровня радиационного загрязнения среды на число проявлений наиболее распространенных опасных заболеваний человека в данном регионе наблюдения. Весьма важной задачей является ввод такого рода информации в компьютерную среду, ее отображение и интерпретация.

Наиболее удобным способом реализации этой задачи представляется применение в комплексе медицинской диагностики интерфейса (сопряжения) с геоинформационными пакетами, что позволит использовать фактографические данные конкретного больного в виде картографических данных (в качестве координат для привязки к карте могут служить данные об области, районе

проживания, месте жительства, работы и т.д.). Такое представление данных больного позволяет визуально определить вклад того или иного показателя в общую клиническую картину. Полученные результаты, как и в предыдущем случае, должны быть занесены в базу знаний.

1.3. Интегрированная среда функционирования современных медицинских систем

В табл. 1 были рассмотрены функции, которые реализуются в различных структурных подразделениях ЛУ (см. рис. 1). Их можно классифицировать следующим образом: 1) функции лечебных отделений; 2) функции приемного отделения; 3) клинико-диагностические функции; 4) функции административно-хозяйственного отделения; 5) функции обучения; 6) функции серверного администрирования.

Чтобы реализовать в рамках информационной системы ЛУ функции различных отделений, необходимо построить локальную сеть ЛУ. На аппаратном уровне такая распределенная система представляет собой совокупность компьютеров или локальных сетей структурных подразделений, соединенных в корпоративную сеть, с возможностью выхода в глобальную сеть.

Структурные подразделения в зависимости от решаемых задач могут использовать различные типы компьютеров и сетевых средств. С точки зрения соотношения цена/производительность и удобства эксплуатации наиболее целесообразно использовать сеть на базе рабочих станций Pentium/ Windows NT и персональных компьютеров. Сеть строится на базе Ethernet 10/100 Мбит/с с поддержкой сетевого протокола TCP/IP. Кроме того, к компьютерам сети подключаются принтеры, сканеры, лабораторное медицинское оборудование и пр.

Наиболее привлекательной и удобной архитектурой для построения такой сети представляется архитектура клиент/сервер. Платформой для реализации распределенной информационной системы ЛУ целесообразно использовать Lotus Notes. Приведем основные достоинства среды Lotus Notes.

Во-первых, данная платформа предназначена для реализации приложений, автоматизирующих делопроизводство любого учреждения. Применительно к медицинскому ЛУ это делопроизводство пациентов ЛУ (проведение больных по

всем аспектам пребывания в ЛУ) и делопроизводство ЛУ как организации (учет больных, бухгалтерский учет, учет лекарственных средств, составление отчетных документов, графиков дежурств врачей и др.). Во-вторых, в Lotus Notes реализован встроенный механизм репликации БД, что обеспечивает обновление БД клиента и сервера, а также встроен механизм блокировок или транзакций, что обеспечивает целостность баз.

Для обеспечения взаимодействия в сети программное обеспечение каждого узла сети состоит из:

- универсального рабочего места пользователя с механизмом автоматической настройки на пользователя;

- программ поддержки распределенной базы данных;

- программ обработки данных, поступающих от лабораторно-медицинского оборудования (при необходимости);

- программ поддержки распределенной базы знаний;

- интеллектуальных систем.

Распределенность БД, баз знаний и возможность взаимодействия различных компонент программного обеспечения узла сети, например БД, экспертных систем и программ обработки данных, обеспечивает взаимодействие различных структурных подразделений ЛУ.

1.4. Типология задач интеллектуализации медицинских систем

Одним из направлений повышения эффективности использования информационно-вычислительных средств является интеллектуализация медицинских систем [7]. Интеллектуализация МС предполагает:

- возможность общения конечного пользователя с системой на формализованном языке предметной области, включая общение аудио- и видеосредствами;

- использование алгоритмов принятия решений на основе исследования проблемы с объяснением этих решений по заданной пользователем неполной или неточной информации;

- возможность манипулирования знаниями;

- использование алгоритмов анализа полученных решений;

использование поисковых процедур и методов решения задач перебора, связанных с поиском и просмотром большого числа вариантов;
применение методов лингвистики в генерации текстов;
возможность динамической пространственной визуализации для представления результатов.

Реализация возможностей интеллектуальных медицинских систем может быть осуществлена с помощью:

систем интеллектуального интерфейса;
систем управления базами знаний;
экспертных систем в различных областях деятельности людей;
мультимедийных баз знаний и данных по областям применения;
средств интеллектуального программирования (языки логического программирования и представления знаний в компьютерах, объектно-ориентированные языки программирования, инструментальные средства разработки программ, системы когнитивной графики);

информационно-обучающих систем (интеллектуальные тренажеры, системы деловых игр, консультационные системы).

Аппаратными средствами интеллектуализации могут быть компьютеры с архитектурой, поддерживающей интеллектуальные программные средства.

Интеллектуализация систем в здравоохранении предполагает решение следующих задач:

разработку средствами мультимедиа электронных медицинских карточек и средств доступа к ним;

создание мультимедийных баз медицинских знаний и данных со средствами удаленного доступа к ним;

разработку диагностических экспертных систем по разделам медицины средствами мультимедиа;

создание баз фармакологических данных со средствами удаленного доступа к ним;

разработку экспертных систем доврачебной помощи для населения со средствами удаленного доступа к ним;

создание тестирующей базы медицинских знаний для выявления и развития способности врача правильно ставить диагноз (на основе мультимедийной базы медицинских данных);

разработку баз знаний со средствами удаленного доступа по технологиям проведения операций и медицинских процедур, а также по использованию медицинской аппаратуры;

создание средств доступа на профессиональном языке к медицинскому оборудованию со встроенными микропроцессорами;

создание экспертных систем и программных средств управления учреждениями здравоохранения.

1.5. Формализация описания предметной области для медицинских интеллектуальных систем

Каждая *предметная область* (ПО) характеризуется тем, что в ней протекает специфическая для нее профессиональная деятельность, которая состоит в решении задач, причем для этого необходимы профессиональные знания, допускающие представление на вербальном уровне.

С целью формализации описания в ПО могут быть выделены **множество объектов ПО**, которые связаны с профессиональной деятельностью в этой ПО, действительность, в которой протекает эта профессиональная деятельность, **знания** об этой действительности (на них базируется профессиональная деятельность) и **соглашения о представлении ПО**, которые позволяют трактовать знания как утверждения о свойствах действительности.

Множество объектов ПО. В множестве объектов ПО существуют подмножества объектов, называемые *величинами*, для каждого из которых существует однозначное отображение **R** этого подмножества на некоторое множество **M**, называемое далее шкалой. Каждый элемент *s* каждой шкалы **M** представляет собой пару (r, m) , где *r* – некоторый математический объект, а *m* – обозначение шкалы.

В различных ПО существуют совокупности объектов, называемые *размерными величинами*. Под размерной величиной **M** понимается такое множество объектов ПО (будем далее называть их *размерные объекты*), для которого определено реализуемое однозначное отображение **R** множества **M** на *размерную шкалу M*. Элементами любой размерной шкалы **M** являются размерные значения – пары (r, m) , где *r* – числа, а *m* – обозначение этой шкалы, называемое размерностью величины. Обозначение каждой размерной шкалы является одновременно и обозначением единицы измерения. С одной размерной величиной

может быть связано несколько размерных шкал с разными единицами измерения. Отображение \mathbf{R} размерной величины \mathbf{M} на размерную шкалу M может быть реализовано с помощью процесса измерения объектов этой размерной величины. Размерные величины возникли в ПО *Физика* (например: длина, измеряемая в метрах, километрах, сантиметрах и т. п.; время, измеряемое в секундах, часах и т. п.; скорость, измеряемая в метрах в секунду, километрах в час и т. п.), однако используются они практически во всех ПО.

В различных ПО существуют совокупности объектов, называемые *скалярными величинами*. Под скалярной величиной понимается такое множество \mathbf{M} объектов ПО (будем далее называть их *скалярные объекты*), для которого определено реализуемое однозначное отображение \mathbf{R} множества \mathbf{M} на (конечную или бесконечную) *скалярную шкалу* M . Элементами любой скалярной шкалы M являются скалярные значения – пары (r, m) , где r – термин ПО или идентификатор, а m – обозначение этой шкалы. Отображение \mathbf{R} скалярной величины \mathbf{M} на скалярную шкалу M может быть реализовано с помощью процесса идентификации объектов скалярной величины со скалярными значениями шкалы M . Этот процесс всегда является субъективным, т. е. выполняется тем же лицом, которое получает результат идентификации. Конечные скалярные шкалы используются во всех ПО, при этом в каждой ПО используются свои скалярные шкалы. Например, в ПО *Медицина* совокупность известных заболеваний (скалярная величина) идентифицируется с множеством медицинских терминов, являющихся названиями этих заболеваний (конечной скалярной шкалы). Примерами *бесконечных скалярных шкал* могут служить шкала идентификаторов D (с обозначением d), шкала фамилий и т. п.

В некоторых ПО существуют величины, объекты которых являются конечными подмножествами некоторой другой величины. Такие величины будем далее называть *величины-множества*, а входящие в них объекты – *объекты-множества*. Шкалы множеств используются, например, в ПО, связанных с учетом кадров в учреждениях. В такой ПО наряду со скалярной шкалой имен всех сотрудников учреждения существует шкала конечных множеств имен сотрудников: значения на этой шкале – множества имен, представляющие коллективы сотрудников различных подразделений этого учреждения.

Во многих ПО существуют величины, объекты которых обладают некоторым числом характеристик – атрибутов, имеющих в качестве значений

объекты некоторых других величин. Такие величины будем далее называть *величины с атрибутами*, а входящие в них объекты – *атрибутированные объекты*.

В некоторых ПО существуют величины, объекты которых являются конечными последовательностями, состоящими из объектов некоторой другой величины. Такие величины будем далее называть *величины-последовательности*, а входящие в них объекты – *объекты-последовательности*. Шкалы последовательностей используются в ПО, где необходимо представлять последовательности объектов (железнодорожные составы, очереди, многошаговые последовательные процессы и т. п.)

Отношения между объектами. Между объектами одной и той же размерной величины определено трехместное отношение *сложение*, которому соответствует трехместное отношение "+" между соответствующими элементами размерной шкалы. Так, если размерная величина – *объем жидкости в сосуде*, соответствующая ей шкала – l , то в отношении сложение находятся объемы жидкости в трех сосудах: объемы жидкости в первых двух сосудах произвольны, жидкость в третьем сосуде получена слиянием жидкости из первых двух сосудов в третий. В отношении "+" входят кортеж $\langle (r_1, l), (r_2, l), (r_1 + r_2, l) \rangle$, где r_1 и r_2 – произвольные положительные вещественные числа; измерив объемы жидкости во всех трех сосудах, находящихся в отношении "сложение", получим, что результаты измерения находятся в отношении "+". Точно так же между объектами одной и той же размерной величины определено трехместное отношение *вычитание*, которому соответствует трехместное отношение "-" между соответствующими элементами размерной шкалы. Аналогично между объектами одной и той же размерной величины определены двухместные отношения *совпадение*, *несовпадение*, *следование за*, *нестрогое следование за*, *предшествование* и *нестрогое предшествование*, которым соответствуют двухместные отношения "=", "≠", ">", "≥", "<", "≤" между элементами соответствующей размерной шкалы.

Между объектами одной и той же скалярной величины существует двухместное отношение *совпадение*, которому соответствует двухместное отношение "=" между элементами соответствующей скалярной шкалы. Так, если скалярная шкала – *моменты времени*, соответствующая ей шкала времен года – {весна, лето, осень, зима}, то в отношении *совпадение* находятся два таких

момента времени, каждый из которых идентифицируется с одним и тем же значением на скалярной шкале, и в отношении "=" входят кортежи $\langle \text{весна, весна} \rangle$, $\langle \text{лето, лето} \rangle$, $\langle \text{осень, осень} \rangle$, $\langle \text{зима, зима} \rangle$; идентифицировав объекты, находящиеся в отношении *совпадение*, получим, что результаты идентификации находятся в отношении "=". Точно так же между объектами одной и той же скалярной величины определено двухместное отношение *несовпадение*, которому соответствует двухместное отношение " \neq ".

Между объектами одной и той же величины-множества определено трехместное отношение *объединение*, которому соответствует трехместное отношение " \cup " между элементами соответствующей шкалы множеств. Так, если величина-множество – *множество людей*, обозначение соответствующей ей шкалы – множество имен, то в отношении *объединение* находятся три множества людей, первые два из которых произвольны, а третье состоит из всех элементов, входящих в первые два множества, и в отношении " \cup " любой кортеж, состоящий из произвольных множеств имен людей и шкалы. Аналогично между объектами одной и той же величины-множества определены двухместные отношения *совпадение*, *несовпадение*, *включение* и *нестрогое включение*, которым соответствуют двухместные отношения "=", " \neq ", "<", " \leq " между элементами соответствующей шкалы множеств.

Между объектами одной и той же величины с атрибутами определены двухместные отношения *совпадение* и *несовпадение*, которым соответствуют двухместные отношения "=" и " \neq " между элементами соответствующей шкалы ролевых фреймов.

Между объектами одной и той же величины-последовательности определено трехместное отношение *сцепление*, которому соответствует трехместное отношение " $\|$ " (*конкатенация*) между элементами соответствующей шкалы последовательностей. Пусть величина–последовательность представляет собой *последовательность людей*, а обозначение соответствующей ей шкалы – последовательность имен. Тогда в отношении *сцепление* находятся три последовательности людей, первые две из которых произвольны, а третья состоит из всех элементов первой последовательности, за которыми следуют элементы второй последовательности в том же порядке, в каком они располагались в исходных последовательностях, и в отношении " $\|$ " входит любой кортеж, состоящий из произвольных последовательностей имен людей и

шкалы. Аналогично между объектами одной и той же величины-последовательности определены двухместные отношения *совпадение, несовпадение, совпадение с началом, совпадение с концом, совпадение с частью, удаление последнего элемента последовательности и удаление первого элемента последовательности*. Им соответствуют свои двухместные отношения.

Знания ПО. Знания ПО могут быть представлены в виде совокупности отношений между следующими специальными терминами предметной области: имена ролей, имена понятий, имена отношений между объектами ситуаций. Каждый кортеж, входящий в любое такое отношение, представляет некоторое утверждение о том, что эти термины, входящие в кортеж, находятся в этом отношении.

Например, в медицинском знании может быть выделено отношение, называемое *область возможных качественных значений медицинского признака*, которое связывает специальные термины, обозначающие медицинские признаки (*наличие боли, периодичность боли* и др.), и термины, обозначающие значения этих признаков (*имеется, ноющая, постоянная* и др.). Примерами утверждений о специальных терминах, находящихся в этом отношении (и представляющих их кортежей, входящих в это отношение), могут быть:

имеется является возможным значением *наличия боли*
(*< наличие боли, имеется >*);

ноющая является возможным значением *характер боли*
(*< характер боли, ноющая >*);

постоянная является возможным значением *периодичность боли*
(*< периодичность боли, постоянная >*).

Другим отношением, которое можно выделить в медицинском знании, является отношение, называемое *область возможных качественных значений фактов*: оно связывает специальные термины, обозначающие факты (*пол, операция по удалению аппендицита* и др.), и термины, обозначающие значения этих фактов (*мужской, не проводилась* и др.). Примерами утверждений (и представляющих их кортежей) могут быть:

мужской является возможным значением факта *пол* (*< пол, мужской >*);

не проводилась является возможным значением факта *операция по удалению аппендицита* (*< операция по удалению аппендицита, не проводилась >*).

Еще одним отношением, которое можно выделить в медицинском знании, является отношение, называемое *клиническая картина заболевания*, которое связывает специальные термины, обозначающие заболевания (*аппендицит* и др.), и термины, обозначающие признаки, входящие в клинические картины этих заболеваний (*боль, локализация боли* и др.). Примерами утверждений (и представляющих их кортежей) могут быть:

наличие боли входит в клиническую картину *аппендицита* (<*аппендицит, наличие боли*>);

локализация боли входит в клиническую картину *аппендицита* (<*аппендицит, локализация боли*>).

Если каждый из кортежей, входящий в некоторое отношение между специальными терминами, состоит из одного элемента, то такое отношение между специальными терминами будем называть *метапонятием*.

Примерами метапонятий ПО *Медицина* могут служить одноместные отношения *значение* (в него входят одноэлементные кортежи *имеется, мужской, не проводилась* и др.), *медицинский признак* (в него входят: *наличие боли, характер боли, периодичность боли, локализация боли* и др.), *факт* (в него входят: *пол, операция по удалению аппендицита* и др.), *заболевание* (в него входят: *аппендицит, холецистит* и др.).

Будем различать функциональные и нефункциональные отношения между специальными терминами. Например, все отношения между специальными терминами для *области возможных качественных значений медицинского признака* являются нефункциональными.

Примером функционального отношения между специальными терминами может служить отношение *нижняя граница возможных размерных значений медицинского признака*, которое связывает названия медицинских признаков с размерными значениями (*температура тела, СОЭ* и др.) и их наименьшие возможные значения. В это отношение могут входить кортежи, представляющие такие утверждения:

наименьшим значением *температуры тела* является 34°C (<*температура тела, 34°C* >);

наименьшим значением *СОЭ* является 1 мм/ч (<*СОЭ, 1 мм/ч* >).

Если отношение между специальными терминами состоит из конечного числа кортежей, то, очевидно, оно может быть представлено перечислением всех

входящих в него кортежей. Если же множество кортежей, входящих в отношение между специальными терминами, бесконечно, то оно не может быть представлено ни в какой форме.

Все кортежи отношения *область возможных качественных значений медицинского признака* связывают два специальных термина: один – из множества объектов *признак с качественными значениями*, а другой – из множества объектов *значение*. Все кортежи отношения *области возможных качественных значений факта* также связывают два специальных термина: один – из множества объектов *факт с качественными значениями*, а другой – из множества объектов *значение*. Каждый кортеж функционального отношения *нижняя граница возможных размерных значений медицинского признака* связывает термин – из множества объектов *медицинский признак с размерными значениями* и термин *ПО нижняя граница возможных размерных значений*.

Соглашения о представлении предметной области. В соглашениях о представлении предметной области используются специальные термины предметной области. Система соглашений о представлении ПО описывает те свойства ролей ПО, понятий ПО и отношений между объектами, а также отношений ПО между специальными терминами, которые не вытекают из предыдущих предположений.

Компонентами модели ПО являются: система шкал, модель действительности, модель знаний и модель соглашений о представлении предметной области, которые определенным образом соответствуют друг другу. Общими для моделей ПО являются способы построения их компонент, а также способ определения соответствия между ними.

2. МАТЕМАТИЧЕСКИЕ МЕТОДЫ ПОСТАНОВКИ ДИАГНОЗА, ИСПОЛЬЗУЕМЫЕ В МЕДИЦИНСКИХ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ

Допустим для простоты изложения, что мы имеем дело с двумя заболеваниями, обозначим их A_1 и A_2 . Предположим, имеется n пациентов с p показателями, характеризующими состояние каждого пациента (анамнез, данные лабораторных и объективных исследований и т. д.). Каждый пациент описывается вектором показателей (признаков) $X = (x_1, x_2, \dots, x_p)^T$, $p > 1$.

Задача постановки диагноза сводится к следующему. По имеющемуся набору признаков $x_{k1}, x_{k2}, \dots, x_{kp}$, полученных у k -го пациента, поставить диагноз A_1 или A_2 .

В зависимости от выбора анализируемых признаков и их физической природы их множества для заболеваний A_1 и A_2 могут быть непересекающимися и пересекающимися. В первом случае задача постановки диагноза сводится к построению разделяющей поверхности L (так называемого диагностического правила или дискриминантной функции). Тогда при предъявлении каждого нового больного с целью постановки диагноза определяется, в какую часть пространства относительно поверхности L попадают значения признаков этого больного. В рассматриваемом примере диагноз может быть поставлен однозначно: у больного заболевание A_1 или A_2 .

Гораздо чаще встречается вторая ситуация, когда множества анализируемых признаков для заболеваний A_1 и A_2 являются пересекающимися. Это означает, что хотя бы один из анализируемых признаков x_i может принимать одно и то же значение для обоих заболеваний. В этом случае задача постановки диагноза также сводится к построению разделяющей поверхности L , но в отличие от первого случая диагноз A_1 или A_2 устанавливается с некоторой вероятностью. Итак, задача диагностирования математически может быть сведена к построению решающего правила (разделяющей поверхности L) и вычислению вероятности постановки правильного диагноза с помощью этого правила. Все сказанное справедливо и для случая, когда нужно диагностировать несколько заболеваний. В этом случае диагностическое правило строится для каждой пары заболеваний.

2.1. Математические методы решения задач медицинской диагностики

Среди математических методов, применяемых для решения задач диагностики заболеваний, можно выделить три группы: вероятностно-статистические, геометрические и логические методы. Наибольшее распространение в медицинской практике получили методы первой группы. Это объясняется наличием развитых теории и программно-алгоритмического обеспечения, а также разработанными принципами интерпретации используемой модели данных и результатов статистических вычислений.

Охарактеризуем вероятностно-статистические методы. В зависимости от способа построения диагностического правила их можно разделить на

теоретико–статистические; эвристические; минимизирующие функцию эмпирического риска.

Теоретико–статистические методы построены на основе теории статистических решений. К ним относятся байесовские методы, метод последовательной процедуры Вальда и др.

Эвристические методы основаны на предположении о том, что каждый класс характеризуется определенным набором параметров – эталоном. Диагностическое правило строится в соответствии с введенной некоторым образом мерой близости к эталону. Примерами эвристических методов диагностики заболеваний являются следующие: метод, основанный на максимизации сходства внутри множества, метод "ближайшего соседа" и др.

Методы, минимизирующие функцию эмпирического риска, строятся на предположении о том, что вид дискриминантной функции известен, а сам алгоритм диагностики характеризуется заданной некоторым образом функцией риска. В этом случае задача построения диагностического правила заключается в определении неизвестных коэффициентов дискриминантной функции. В зависимости от выбранной функции эмпирического риска и от метода оценивания неизвестных коэффициентов дискриминантной функции существует большое число различных алгоритмов. Совокупность алгоритмов, порождающих на основании предположений и выборки конкретное правило классификации, называют дискриминантным анализом (ДА) [8]. Рассмотрим далее базовые понятия ДА.

2.2. Классификация при наличии обучающих выборок (дискриминантный анализ)

В общем виде задача различения (дискриминации) формулируется следующим образом. Пусть результатом наблюдения над объектом является реализация p -мерного случайного вектора $X = (x_1, x_2, \dots, x_p)^T$. Требуется установить правило, согласно которому по наблюденному значению вектора X объект относится к одной из возможных совокупностей $p_i, i = 1, 2, \dots, g$. Для построения правила дискриминации (решающего правила) все выборочное пространство R значений вектора X разбивается на области $R_i, i = 1, 2, \dots, g$ так, что при попадании X в R_i объект относят к совокупности p_i .

Предполагается, что случайная выборка представляет собой последовательность независимых пар наблюдений вида (X_i, y_i) , $i = 1, \dots, n$, где $P\{y_i = j\} = p_j$, $j = 1, \dots, g$ – номер класса, к которому принадлежит наблюдение X_i , p_j – неизвестная вероятность того, что X будет извлечено из j -го класса. Число классов g заранее известно исследователю, известно также, что все X_i принадлежат одному и тому же пространству наблюдений; X_i – такие, что $y_i = j$, одинаково распределены с неизвестной функцией распределения F_j . Такую выборку принято называть обучающей. Число $y_i = j$ в выборке будем называть объемом выборки из j -го класса и обозначать n_j . При построении правила классификации используются два вида информации: предположения о свойствах распределения X_i внутри классов и обучающая выборка.

Обычно в задаче различения переходят от вектора признаков, характеризующих объект, к линейной функции от них, дискриминантной функции – гиперплоскости, наилучшим образом разделяющей совокупности выборочных точек. Наиболее изучен случай, когда известно, что распределение вектора признаков в каждой совокупности нормальное, но нет информации о параметрах этих распределений. Здесь естественно заменить неизвестные параметры распределения в дискриминантной функции их наилучшими оценками. Непараметрические методы дискриминации не требуют знания точного функционального вида распределения и позволяют решать задачи на основе незначительной априорной информации.

Процедура дискриминантного анализа разделяется на два этапа. На первом этапе решаются задачи, связанные с формированием обучающей выборки (отбор информативных признаков, проверка выполнения основных допущений метода ДА, отбор наблюдений). Это необходимо для нахождения такого представления наблюдений в дискриминантном пространстве, которое обеспечит в некотором смысле наилучшее разделение g классов. На втором этапе ДА производится классификация, т. е. отнесение каждого нового наблюдения, чья принадлежность к какому-либо из классов априори неизвестна, к одному из g заданных классов. Процесс классификации сопровождается оценкой ошибки классификации.

В качестве решающего правила широко используется линейная дискриминантная функция Фишера (ЛДФ) [8]: наблюдение принадлежит тому классу, для которого значение ЛДФ максимально. Выбор этого метода

объясняется рядом преимуществ ЛДФ: ЛДФ устойчива к нарушению ряда допущений ДА [8]; величина стандартизованных коэффициентов ЛДФ позволяет понять и интерпретировать различия между классами; проекции обучающей выборки на ЛДФ могут быть изучены графически; алгоритмы вычисления ЛДФ реализованы во многих известных статистических пакетах программ [9]. Кроме того, для принятия решения о классификации наблюдения оцениваются апостериорные вероятности отнесения его к тому или иному классу. В этом случае прогностическое правило формулируется следующим образом: наблюдение принадлежит тому классу, для которого значение апостериорной вероятности максимально.

Библиотека БГУИР

ЛИТЕРАТУРА

1. Айламазян А. К., Осипов Г. С. Проблемы создания интегрированной среды поддержки лечебно-диагностического процесса / Информационные технологии, 1997, № 10.
2. Статические и динамические экспертные системы: Учеб. пособие / Э. В. Попов и др. М.: Финансы и статистика, 1996.
3. Попов Э. В. Экспертные системы: Решение неформализованных задач в диалоге с ЭВМ. М.: Наука, 1987.
4. Гублер Е. В. Вычислительные методы анализа и распознавания патологических процессов. Л.: Медицина, 1984.
5. Осипов Г. С. Приобретение знаний интеллектуальными системами: основы теории и технологии. М.: Наука, 1997.
6. Голенков В. В. Графодинамические методы и средства параллельной асинхронной переработки информации в интеллектуальных системах. Мн.: БГУИР, 1996.
7. Кобринский Б.А. Системы искусственного интеллекта в медицине: состояние, проблемы и перспективы // Новости искусственного интеллекта. 1995. №2.
8. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: классификация и снижение размерности. М.: Финансы и статистика, 1989.
9. Боровиков В. П., Боровиков И. П. STATISTICA – Статистический анализ и обработка данных в среде Windows. М.,1997.