

На правах рукописи

Москаленко Филипп Михайлович

**МЕТОДЫ РЕШЕНИЯ ЗАДАЧИ МЕДИЦИНСКОЙ ДИАГНОСТИКИ
НА ОСНОВЕ МАТЕМАТИЧЕСКОЙ МОДЕЛИ
ПРЕДМЕТНОЙ ОБЛАСТИ**

05.13.18 – математическое моделирование, численные методы
и комплексы программ

Автореферат

диссертации на соискание ученой степени
кандидата технических наук



Владивосток

2010

Работа выполнена в лаборатории интеллектуальных систем
Института автоматики и процессов управления ДВО РАН

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

Научный консультант: доктор медицинских наук
Черняховская Мери Юзефовна

Официальные оппоненты: доктор технических наук, профессор
Девятисильный Александр Сергеевич

доктор технических наук, профессор
Кулешов Евгений Львович

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

Защита состоится «11» февраля 2011 года в 10 часов на заседании
диссертационного совета Д.005.007.01 в Институте автоматики и процессов
управления ДВО РАН по адресу: 690041, г. Владивосток, ул. Радио, 5.

С диссертацией можно ознакомиться в библиотеке Института автоматики
и процессов управления ДВО РАН.

Автореферат разослан «28» декабря 2010 г.

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


А.В. Лебедев

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы. Системы медицинской диагностики являются одним из приложений методов искусственного интеллекта. Их разработка имеет целью помочь врачу избежать собственных ошибок. Задачей таких систем является определение заболевания, которым болен пациент, на основе данных о его наблюдениях и построение объяснения принятого решения.

Значительный вклад в разработку систем медицинской диагностики для ЭВМ внесли отечественные и зарубежные ученые: Н.Н. Амосов, И.П. Быховский, А.А. Вишневский, И.М. Гельфанд, Е.В. Гублер, А.С. Клещев, Б.А. Кобринский, Г.А. Хай, М.Ю. Черняховская, В. Chandrasekaran, С.А. Kulikowski, R.S. Ledley, L.B. Lusted, F. Mizoguchi., S.G. Pauker, E.H. Shortliffe, P. Szolovits и другие.

Были разработаны модели медицинской диагностики, описывающие знания в форме правил, фреймов, патофизиологических моделей, а также модели онтологий медицинской диагностики; предложены методы решения задачи диагностики, которые либо моделируют рассуждения врача, либо обрабатывают заданные экспертами правила, либо проверяют соответствие состояния больного и клинических картин заболеваний, описанных врачом; разработаны методы оптимизации алгоритмов диагностики, состоящие в использовании извлекаемой из баз знаний дополнительной информации или в использовании специальных сред разработки и техник программирования; некоторые системы медицинской диагностики созданы на базе сетевых технологий и предоставляют доступ к системе через сеть Интернет; наиболее широко в практической медицине применяются узконаправленные системы, способные проводить диагностику одного или небольшого числа заболеваний в какой-то области.

Однако модели знаний, лежащие в основе современных экспертных систем медицинской диагностики, все еще заметно проще реальных представлений врачей, большинство таких систем имеет локальную реализацию, а время работы многих систем всё ещё чрезмерно велико.

Таким образом, актуальной задачей является разработка методов решения задачи медицинской диагностики, основанных на математической модели онтологии, приближенной к реальным представлениям в области медицины, в терминах которой знания представлены в форме, понятной врачам, и которые позволяют определять не только диагноз пациента, но и причину каждого заболевания, а также объяснять изменение во времени наблюдаемых значений признаков с учетом индивидуальных особенностей пациента и происшедших с ним событий. При этом система, основанная на таких методах, должна быть способна ставить диагноз в приемлемое для врача время и предоставлять удаленный доступ многим пользователям-врачам.

Целью диссертационной работы является разработка и исследование математической модели онтологии, методов и комплекса программ для решения задачи медицинской диагностики на многопроцессорной вычислительной системе.

Для достижения этой цели необходимо решить следующие **задачи**:

- 1) разработать математическую модель онтологии медицинской диагностики, отвечающую современным представлениям экспертов в этой области;
- 2) сформулировать задачу медицинской диагностики в терминах этой онтологии, разработать методы её решения и реализующий их алгоритм для многопроцессорной вычислительной системы;
- 3) разработать концепцию и методы реализации комплекса программ для решения задачи медицинской диагностики с использованием многопроцессорной вычислительной системы в качестве сервера приложений;
- 4) провести экспериментальное исследование вычислительной сложности реализованного алгоритма решения задачи медицинской диагностики и исследовать практическую применимость разработанного комплекса программ.

Методы исследования. Для решения указанных задач применялись элементы теории множеств, математической логики, теории формальных языков, методы разработки онтологий и построения их моделей, методы системного, объектно-ориентированного, параллельного программирования, методы разработки пользовательских интерфейсов.

Научная новизна работы состоит в следующем:

- впервые разработаны онтология медицинской диагностики и её математическая модель, отличающиеся от известных максимальной приближенностью к реальным представлениям экспертов в области медицинской диагностики в нашей стране;
- впервые разработаны метод и алгоритм решения задачи медицинской диагностики, основанные на математической модели нетривиальной онтологии, а так же доказаны теоремы об остановке и корректности алгоритма;
- впервые сформулирована концепция сетевого ресурса, предназначенного для обеспечения процесса согласованного решения комплекса задач по хранению, управлению и обработке данных и знаний в области медицинской диагностики острых заболеваний, с возможностью удаленного доступа.

Практическая ценность работы состоит в следующем:

- реализован алгоритм медицинской диагностики, способный работать в последовательном и параллельном режимах, основанный на математической модели нетривиальной онтологии и производящий поиск соответствия истории болезни пациента и картин заболеваний, описанных врачами;
- разработан комплекс программ (с применением платформы облачных вычислений *Многоцелевой банк знаний*) для решения задачи медицинской диагностики, способный проводить диагностику острых заболеваний в любой области медицины в приемлемое время и при этом выдавать понятное врачу объяснение полученного результата.

Реализация результатов работы. Результаты работы использованы в учебном процессе:

- во Владивостокском государственном медицинском университете – материалы второй и пятой глав используются для знакомства студентов с фор-

мальным представлением острых заболеваний глаза и его придаточного аппарата в курсах лекций по офтальмологии, материалы четвертой главы используются для практики в постановке диагнозов в области офтальмологии;

- в Дальневосточном государственном университете – материалы диссертации используются в учебном процессе при чтении курса лекций по СД «Системы искусственного интеллекта» студентам специальности 010503.65 «Математическое обеспечение и администрирование информационных систем».

Результаты работы использованы в научной работе и исследованиях сотрудников кафедры ПО ЭВМ в Институте математики и компьютерных наук ДВФУ и сотрудников лаборатории интеллектуальных систем ИАПУ ДВО РАН.

Положения, выносимые на защиту:

- модель онтологии медицинской диагностики, максимально приближенная к реальным представлениям экспертов;
- метод решения частной задачи медицинской диагностики и реализующий его корректный алгоритм;
- доступный через сеть Интернет комплекс программ, позволяющий проводить медицинскую диагностику за приемлемое время, а также редактировать базы знаний, наблюдений и историй болезни.

Обоснованность и достоверность результатов работы обеспечиваются применением теоретических методов разработки интеллектуальных систем и подтверждены результатами экспериментальных исследований.

Апробация работы. Основные научные и практические результаты работы докладывались и обсуждались на следующих международных и российских конференциях и семинарах: втором Международном научно-практическом семинаре «Высокопроизводительные параллельные вычисления на кластерных системах» (Нижний Новгород, 2002), Дальневосточной математической школе-семинаре имени академика Е.В. Золотова (Владивосток, 2002, 2003, 2004, 2006, 2008; Хабаровск, 2005), открытым Дальневосточном конкурсе программных средств студентов, аспирантов и молодых специалистов «Программист» (Владивосток, 2003, 2004, 2006, 2009), II и III международных конференциях «Параллельные вычисления и задачи управления» (Москва, 2004, 2006), XIII международной конференции «Знание-Диалог-Решение» (Болгария, Варна, 2007), научной сессии МИФИ (Москва, 2008), конференции «Системный анализ в медицине» (Благовещенск, 2009), Первой Российско-Тихоокеанской конференции в области компьютерных технологий и приложений (Владивосток, 2010).

Публикация результатов работы. По материалам диссертации опубликованы 32 печатные работы, в том числе: 5 статей в журналах, входящих в Перечень ВАК РФ, 7 статей в других российских и иностранных журналах, 3 препринта, 17 тезисов и трудов конференций.

Структура и объём работы. Диссертационная работа состоит из введения, пяти глав, заключения, списка литературы из 128 наименований, и трёх приложений. Основная часть работы изложена на 139 страницах, содержит 5 таблиц и 12 рисунков.

СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обосновывается актуальность выполненных в диссертации исследований, формулируется цель работы, рассматривается научная и практическая значимость результатов, приводятся сведения об апробации и реализации основных положений диссертации.

Первая глава представляет собой обзор литературы, в котором рассматриваются существующие системы медицинской диагностики, основанные на знаниях, и их составляющие. При этом особое внимание уделяется таким характеристикам систем, как: использование моделей знаний, основанных на сложных онтологиях; присутствие элементов оптимизации алгоритма и/или базы знаний; использование многопроцессорных вычислительных систем; возможность расширения базы знаний экспертами.

Во **второй главе** приводится описание онтологии (и её модели) предметной области «медицинская диагностика», которая учитывает взаимодействие причинно-следственных отношений различных типов. Данная онтология близка к реальным представлениям медицины в нашей стране и описывает сочетанную и осложненную патологии, динамику патологических процессов во времени, а также воздействие лечебных мероприятий и других событий на проявления заболеваний. Математическая модель онтологии медицинской диагностики представляется в виде небогатой системы логических соотношений с параметрами на языке прикладной логики и включает определения терминов модели знаний (параметры), определения терминов модели действительности (неизвестные), а также ограничения целостности неизвестных и параметров и соотношения между ними. Термины и онтологические соглашения в модели объединены в три раздела: базовые понятия и онтологические соглашения, описание причинно-следственных связей и описание причин значений признаков и заболеваний.

Первая часть модели описывает базовые понятия и онтологические соглашения знаний и действительности предметной области. К базовым терминам знаний относятся: *признаки, события, особенности, наблюдения, множества значений, условия, необходимое условие, заболевания, периоды динамики, интервал*. К базовым терминам действительности относятся: *моменты, наблюдавшиеся особенности, выполнено, диагноз, развитие, интервалы развития признака*. Также в этой части модели описываются онтологические соглашения, задающие ограничения на значения базовых терминов знаний и действительности. Например, соглашение о том, что названия всех признаков, событий и особенностей различны, выглядит так:

$$\begin{aligned} \text{признаки} \cap \text{события} &= \emptyset \ \& \ \text{особенности} \cap \text{события} = \emptyset \ \& \\ \& \ \text{признаки} \cap \text{особенности} &= \emptyset \end{aligned}$$

Вторая часть модели содержит термины знаний и действительности, описывающие различные причинно-следственные связи, их ограничения целост-

ности, а также онтологические соглашения о соответствии между знаниями и действительностью. Они сгруппированы в семь разделов: общие термины и онтологические соглашения, описание нормальных реакций, описание реакций на воздействие событий, описание клинических проявлений, описание клинических проявлений, измененных воздействием событий, описание этиологий, описание осложнений. Например, описание нормальных реакций задаётся с помощью пяти предложений на языке прикладной логики:

- Термин *знания о нормальных реакциях* это множество структурных значений с атрибутами *следствие*, *варианты*, *воздействующие факторы*, каждое из которых описывает знания о конкретной нормальной реакции. Значением первого атрибута является имя признака, второго – множество вариантов нормальной реакции для этого признака, третьего – множество особенностей.

знания о нормальных реакциях \equiv
 \equiv (*следствие* \rightarrow *признаки*, *варианты* \rightarrow { } *варианты нормы*,
воздействующие факторы \rightarrow { } *особенности*)

- Термин *знаний варианты нормы* это множество структурных значений с атрибутами *область значений следствия*, *условия на воздействующие факторы*, каждое из которых описывает знания о конкретном варианте нормальной реакции. Значением первого атрибута является множество значений признака в этом варианте, а второго – условие.

варианты нормы \equiv (*область значений следствия* \rightarrow *множества значений*,
условие на воздействующие факторы \rightarrow *условия*)

- В знаниях о нормальных реакциях в любом варианте нормы *область значений следствия* – это собственное подмножество возможных значений признака, являющегося следствием этой нормальной реакции.

(знания о нормальной реакции: *знания о нормальных реакциях*)
 (вариант: *варианты*(знания о нормальной реакции))

область значений следствия(вариант) \subset
 \subset *возможные значения*(*следствие*(знания о нормальной реакции))

- Термин *действительности нормальные реакции* это множество структурных значений с атрибутами *следствие* и *вариант*, каждое из которых описывает нормальную реакцию, происходящую в ситуации. Значением первого атрибута является имя признака, а второго – вариант нормы.

нормальные реакции \equiv (*следствие* \rightarrow *признаки*,
вариант \rightarrow *варианты нормы*)

- Если в ситуации присутствует некоторая нормальная реакция, то в множестве *знания о нормальных реакциях* присутствует такой элемент, у которого следствие совпадает со следствием нормальной реакции, в множестве вариантов нормы этого элемента содержится вариант нормы из нормальной реакции и для него выполнено условие на воздействующие факторы.

(нормальная реакция: *нормальные реакции*)
 (∨ (знания о нормальной реакции: *знания о нормальных реакциях*)
следствие(знания о нормальной реакции) =
 = *следствие*(нормальная реакция) &
 & *вариант*(нормальная реакция) ∈
 ∈ *варианты*(знания о нормальной реакции) &
 & *выполнено*(условие на воздействующие
факторы(*вариант*(нормальная реакция))))

В третьей части модели онтологии определяются термины, описывающие причины значений признаков на интервалах их развития, а также термины, описывающие причины заболеваний. При описании причин значений признака на интервале его развития приводятся соотношения, устанавливающие, какие из причинно-следственных связей, действующих на некотором интервале времени, связанном с наблюдаемым признаком, определяют значения этого признака. Для этого определяется множество всех возможных причинных закономерностей, которые могли бы регулировать значения рассматриваемого признака для всех моментов времени, попадающих в этот интервал. Вводится понятие приоритета причины значений признака. Приоритет зависит от модальности причинно-следственной связи («необходимость» приоритетнее «возможности») и от того, к какому классу она принадлежит (порядок увеличения приоритета: нормальная реакция << реакция на воздействие события << клиническое проявление << клиническое проявление, изменённое воздействием события << клиническое проявление заболевания-осложнения << клиническое проявление заболевания-осложнения, изменённое воздействием события). Причинно-следственная связь, определяющая значения признака на этом интервале, имеет максимальный приоритет в множестве возможных причин.

При описании причин заболеваний приводятся соотношения, устанавливающие, какая причинно-следственная связь (из множества этиологий и осложнений) является причиной заболевания из диагноза. Для этого определяется множество всех возможных причинных закономерностей, которые могли бы быть причиной рассматриваемого заболевания. Вводится понятие приоритета причины заболевания, который зависит от модальности причинно-следственной связи и от того, когда она началась (более ранняя причина имеет больший приоритет). Причинно-следственная связь, являющаяся причиной заболевания, имеет максимальный приоритет в множестве возможных причин.

Третья глава содержит постановку общей задачи медицинской диагностики (прямой и обратной) в терминах математической модели её онтологии, постановку частной обратной задачи и описание методов её решения: последовательного, модифицированного и параллельного – для многопроцессорной вычислительной системы. Здесь же приводятся формулировки теорем об

остановке и корректности алгоритма, реализующего эти методы, эквивалентности последовательной версии и модифицированной параллельной версии, а также даются сведения о реализации модифицированной параллельной версии.

Задача медицинской диагностики, наиболее часто решаемая врачами в их практике, состоит в определении всех возможных альтернативных *диагнозов* больного на основе знаний предметной области и данных его обследования, к которым относят *значения признаков* (в моменты их наблюдения), *значения анатомо-физиологических особенностей* (постоянные во времени) и *значения произошедших событий* (в моменты, когда они происходили). В каждый *диагноз* могут входить несколько *заболеваний*, протекающих одновременно или последовательно. Считается, что все *наблюдения* протекают на определённом временном отрезке, называемом *периодом наблюдения*, на котором моменты времени измеряются вещественными числами, каждое из которых есть число часов, прошедших с момента начала наблюдений (нуля часов). Считается, что до этого момента, наблюдения признаков отсутствуют, а заболевания не протекают. Такая задача относится к классу *обратных задач*. Ей соответствует *прямая задача*, состоящая в определении возможных *состояний* пациента (значений наблюдаемых признаков) в те или иные *моменты времени* на основе известного *диагноза*, *значений особенностей* и происходящих *событий*.

В терминах модели онтологии математическая постановка обратной задачи может быть сформулирована следующим образом:

Дано:

- медицинские знания, удовлетворяющие ограничениям целостности знаний:
 - *знания о наблюдениях (признаках, событиях, анатомо-физиологических особенностях* пациента) и их *возможных значениях*;
 - *знания о заболеваниях и длительностях их периодов развития*;
 - *знания о возможных причинах заболеваний – знания об этиологиях и знания об осложнениях*;
 - *знания о значениях признаков при отсутствии заболеваний – знания о нормальных реакциях и знания о реакциях на воздействие событий*;
 - *знания о причинно-следственных связях между заболеваниями и наблюдениями – знания о клинических проявлениях и знания о клинических проявлениях, изменённых воздействием событий*;
- результаты наблюдений больного:
 - *наблюдавшиеся анатомо-физиологические особенности* и результаты их наблюдения (*значения*),
 - *наблюдавшиеся признаки, моменты их наблюдения и их значения* в эти моменты,
 - *произошедшие события, моменты, в которые они происходили, и их значения* в эти моменты.

Найти:

- все альтернативные *диагнозы* пациента (один диагноз – множество *заболеваний*, которыми болен пациент на протяжении *периода наблюдений*), и:
 - *объяснение* каждого диагноза – *причину* каждого *заболевания* в *диагнозе* (*вариант этиологии* или *вариант осложнения*),
 - *время начала* и *время конца* каждого периода развития каждого *заболевания*,
 - *причины* всех *наблюдаемых значений признаков* (*нормальные реакции, реакции на воздействие события, клинические проявления заболеваний и клинические проявления, изменённые воздействием событий*, возможно, с указанием номера *периода динамики* и/или *причины-события*) при каждом *диагнозе*.

Связи между тем, что дано, и тем, что требуется найти, определяются моделью онтологии и базой знаний.

В связи с тем, что применяемая модель онтологии учитывает большое число связей между процессами, протекающими в организме больного, ожидается, что метод решения сформулированной выше задачи медицинской диагностики, анализирующий все эти связи, будет иметь высокую вычислительную сложность. Поскольку в данной работе рассматривается детерминированная модель, то основным видом зависимости между патологическими процессами является их синхронизация по времени. Такие зависимости возникают, когда больной болен одновременно несколькими заболеваниями (поскольку клинические проявления каждого из этих заболеваний синхронизированы относительно его начала), при осложнениях одних заболеваний другими (поскольку клинические проявления осложнения синхронизированы относительно начала этого осложнения), при наличии у заболевания более чем одного периода развития (поскольку в каждом периоде развития его клинические проявления синхронизированы относительно начала и конца этого периода). Когда моменты, относительно которых синхронизированы разные процессы, неизвестны, вычислительная сложность задачи диагностики значительно возрастает из-за дополнительного перебора гипотез о значениях этих моментов.

В случае острых заболеваний (за исключением тяжёлых травм) на начальном (т.е. первом) этапе их развития практически важно поставить диагноз больному как можно скорее; при этом осложнения еще, как правило, не развились. В этой ситуации известны все моменты времени, относительно которых синхронизированы разные процессы: начало заболевания и моменты, в которые происходили различные события (например, связанные с экстренными лечебными мероприятиями). Поэтому процессы, соответствующие изменению значений разных признаков, могут обрабатываться независимо друг от друга.

В этом случае может быть сформулирована частная практически полезная обратная задача медицинской диагностики (и соответствующая ей част-

ная прямая задача), которая получается из поставленной выше, если на модель онтологии наложить следующие ограничения:

- *диагноз* пациента может включать не более одного *заболевания* (на протяжении всего *периода наблюдений* пациент болен одним *заболеванием* или *здоров*);
- в качестве причин заболеваний не рассматриваются *осложнения* одного *заболевания* другим;
- каждое *заболевание* имеет единственный *период развития*;
- если пациент болен некоторым *заболеванием*, то момент *начала* этого *заболевания* известен.

Благодаря тому, что при решении такой обратной задачи процессы, соответствующие изменению значений разных признаков, могут обрабатываться независимо друг от друга, для повышения эффективности метода решения задачи диагностики может быть выполнено его распараллеливание, что позволит исполнять реализованный алгоритм на многопроцессорной вычислительной системе (МВС).

Кроме того, при описанных выше условиях сокращается количество анализируемых причинно-следственных связей (ПСС) при поиске причин значений признака (если признак входит в клиническую картину заболевания, то рассматриваются только клинические проявления и клинические проявления, измененные воздействием события; если же признак не входит в клиническую картину заболевания, то рассматриваются только нормальные реакции и реакции на воздействие события).

Следующие допущения являются естественным дополнением предыдущих:

- если пациент болен, то это *заболевание* протекает на протяжении всего *периода наблюдений* (т.е. никакой *признак* и никакое *событие* не имеют значений после окончания *заболевания*);
- *наблюдения признаков* производятся не *ранее начала заболевания* из *диагноза*, тогда как *события* могут происходить до *начала заболевания* – в отрицательные *моменты* времени;
- при наблюдении *пациента* известны значения всех его *анатомо-физиологических особенностей*.

Естественным методом решения частной обратной задачи медицинской диагностики является перебор всех возможных значений выходных данных (диагнозов), при котором для каждого заболевания выполняется решение прямой задачи – построение всех возможных схем развития ПСС и поиск среди них тех, которым соответствуют все наблюдаемые значения признаков пациента.

Результатом решения частной обратной задачи является:

- либо подтверждённая гипотеза о том, что пациент здоров, с объяснением (указанием причин наблюдаемых значений признаков);

- либо несколько взаимоисключающих подтверждённых гипотез о диагнозе, каждая из которых представляет собой одно заболевание, которым болен пациент на всём периоде наблюдений, вместе с объяснением (указанием причин наблюдаемых значений признаков и, возможно, причины самого заболевания);
- либо отказ от диагноза (пациент болен неизвестным заболеванием или введённые результаты наблюдений пациента противоречат базе знаний – ошибка в данных или знаниях).

Перебор гипотез о диагнозе начинается с гипотезы о том, что пациент здоров. В случае её подтверждения прочие гипотезы не рассматриваются.

При рассмотрении каждой гипотезы наблюдаемые признаки обрабатываются независимо друг от друга. При работе с каждым признаком ищется причина каждого его наблюдаемого значения. Это выполняется последовательно, от самого раннего по времени наблюдения значения до самого позднего. Для поиска причины очередного значения признака необходимо знать, какие связи присутствуют между процессами, протекающими в организме пациента, в момент наблюдения, и найти среди них ту, которая является причиной наблюдаемого значения, т.е. для которой наблюдаемое значение принадлежит области значения следствия. Для этого на основе информации из базы знаний (с учётом значений событий и анатомо-физиологических особенностей) строятся все возможные схемы развития во времени ПСС, каждая из которых протекает согласно одному из возможных вариантов её развития. Любая ПСС начинается в определённый момент времени в соответствии с одним из вариантов своего развития и протекает определённое время:

- реакция на воздействие события (РВС) или клиническое проявление, изменённое воздействием события (КПИВС), начинается в момент, когда происходит событие (при выполнении необходимого условия этой ПСС, и для её варианта – при выполнении условия на воздействующие факторы и вхождении значения события в множество допустимых значений), а длительность протекания этой ПСС определяется длительностью периодов динамики используемого варианта этой ПСС;
- нормальная реакция (НР) или клиническое проявление (КП) начинается в момент начала наблюдений (0 часов), что также является моментом начала заболевания (при выполнении условия на воздействующие факторы для варианта этой ПСС, а для КП – также при выполнении необходимого условия) и протекает на протяжении всего периода наблюдения (для варианта КП это так, потому что сделано предположение о том, что пациент наблюдается только во время протекания заболевания).

ПСС, развивающиеся в соответствии с одним из возможных вариантов, накапливаются (добавляются в момент гипотетического начала их действия и удаляются по окончании действия) в множестве гипотез о развитии вариантов

причинно-следственных связей (МГРВПСС). В нём все ПСС делятся на два подмножества: с высоким (МГРВПССвп) и с низким приоритетом (МГРВПССнп). ПСС в каждом подмножестве хранятся в отсортированном виде, критерий – момент начала действия варианта ПСС.

Таким образом, в каждый момент наблюдения некоторого признака в МГРВПСС ищется тот вариант, который является причиной рассматриваемого значения, то есть:

- который имеет наивысший приоритет среди других в МГРВПСС,
- в множество значений следствия которого (для периода динамики, протекающего в момент наблюдения) входит это наблюдаемое значение признака.

Поскольку алгоритм, реализующий метод решения частной обратной задачи медицинской диагностики, при достаточно большом наполнении базы знаний вынужден выполнять переборы высокой вычислительной сложности, то время его работы при таких условиях (теоретически) велико. В связи с этим встаёт вопрос о разработке оптимизированных методов и реализующих их алгоритмов с целью уменьшения времени, затрачиваемого на решение этой задачи. Решить задачу оптимизации можно двумя путями:

- за счет сокращения переборов, выполняемых алгоритмом, при использовании некоторых дополнительно извлекаемых знаний;
- за счет распараллеливания алгоритма для его выполнения на МВС.

Для выполнения первой оптимизации сокращается перебор гипотез о диагнозе, то есть возможных заболеваний. Очевидно, что определённые значения признаков могут наблюдаться только при протекании у пациента некоторых заболеваний, количество которых может быть либо сопоставимо с количеством заболеваний в базе знаний, либо существенно меньше его. Таким образом, для сокращения первого перебора, необходимо по базе знаний для каждого значения каждого признака определить, при каких заболеваниях это значение может наблюдаться (максимум – все заболевания, минимум – ни одного). На основе этой информации и наблюдений пациента при начале диагностики становится возможным получить для каждого наблюдаемого значения множество заболеваний, при которых оно может встретиться, и в качестве множества гипотез рассматривать пересечение всех таких множеств.

Для распараллеливания в алгоритме решения задачи диагностики необходимо выделить вычислительную часть, несколько экземпляров которой можно запускать одновременно на вычислительных узлах МВС для независимой параллельной обработки наблюдаемых признаков, и управляющую часть. Исследование алгоритма медицинской диагностики показывает, что в нем возможно распараллеливание на одном из двух уровней: параллельно исследуются гипотезы о диагнозе или параллельно исследуются наблюдаемые признаки для каждой гипотезы. Ввиду того, что обработка одних гипотез о диагнозе может занять существенно больше времени, чем обработка других,

распараллеливание производится на втором уровне. В этом случае в функции управляющей части входит формирование множества гипотез о диагнозе для проверки, формирование заданий вида <заболевание; признак> для вычислительных частей, их раздачу и приём результатов, формирование итогового результата.

Реализация параллельного модифицированного алгоритма медицинской диагностики выполнена на языке ANSI C++ с применением библиотеки функций параллельного программирования MPI, что позволяет компилировать и запускать её на различных платформах.

В **четвёртой главе** описана концепция банка знаний по медицинской диагностике и приводится описание его реализации в рамках платформы облачных вычислений *Многоцелевой банк знаний*.

Банк знаний по медицинской диагностике – это комплекс программ, который обеспечивает процесс согласованного решения задач по сбору, хранению, управлению и обработке данных и знаний в области медицинской диагностики и является объединением всей этой информации в единый ресурс с возможностью удаленного доступа к нему многим пользователям.

Архитектура банка включает следующие компоненты:

- информационное наполнение (ИН), хранящееся на сервере банка знаний и содержащее знания (о наблюдениях и заболеваниях) и данные (истории болезни пациентов), сформированные согласно модели;
- программное наполнение (ПН), ориентированное на интеллектуальную поддержку пользователей банка, разработанное в соответствии с технологией «клиент-сервер» (в совокупности с ИН образуется «трехзвенная» архитектура) и включающее:
 - средства по редактированию данных из ИН (баз знаний о наблюдениях и заболеваниях и базу историй болезни), выполняющие также роль подсистемы доверия, так как они обеспечивают возможность просмотра базы знаний,
 - средства по обработке данных из ИН: оптимизатор базы знаний о заболеваниях и интеллектуальный решатель задачи медицинской диагностики,
 - административную подсистему.

Кроме того, отметим следующие особенности архитектуры:

- ИН банка состоит из разделов, каждый из которых содержит знания и данные конкретной области медицинской диагностики (офтальмология, иммунология и т.п.);
- поскольку алгоритм диагностики предназначен для реализации на МВС, то часть средств ПН и все компоненты ИН находятся на сервере банка знаний, а подсистема, непосредственно выполняющая диагностику – на МВС.

Пользователи банка знаний делятся на следующие группы:

- эксперты, которые формируют базы знаний о наблюдениях и о заболеваниях;
- гости банка, которые знакомятся с применяемыми базами знаний;

- врачи, которые:
 - формируют истории болезни пациентов,
 - пользуются подсистемой медицинской диагностики;
- администратор, который:
 - выполняет функции главного эксперта банка – на основе персональных баз знаний экспертов формирует общую базу знаний о заболеваниях для каждого раздела банка и выполняет её оптимизацию (выделение неявной информации о связи признаков и заболеваний),
 - осуществляет управление полномочиями пользователей.

Архитектура компонент сервер-части банка знаний по медицинской диагностике представлена на рис. 1, где присутствуют следующие обозначения:

- прямоугольниками обозначены компоненты ИН и ПН;
- стрелки, имеющие изгибы под прямым углом, обозначают передачу информации между компонентами ИН, ПН и МВС;
- изогнутая стрелка обозначает вызов программы диагностики.



Рис.1. Архитектура сервер-части банка знаний по медицинской диагностике

Реализация прототипа банка знаний по медицинской диагностике выполнена с использованием разработанной в ИАПУ ДВО РАН платформы Многоцелевого Банка Знаний (МБкЗ), предназначенной для поддержки жизненного цикла совместимых систем обработки информации. При этом программное наполнение разработанного специализированного банка знаний (СБкЗ) по медицинской диагностике доступно пользователям через сеть Интернет и включает в себя:

- редактор ИРУО, являющийся частью МБкЗ, который используется для редактирования на языке ИРУО моделей онтологий и целевой информации, формируемой по этим моделям – баз знаний и историй болезни;
- информационно-административную систему (ИАС), также входящую в состав МБкЗ, с помощью которой выполняется управление правами пользователей;
- решатель задачи медицинской диагностики;
- преобразователь базы знаний о заболеваниях (публикатор).

Доступ к МВС предоставляется серверу МБкЗ. На нём работает слабо-функциональная серверная часть решателя, реализованная на языке Java, которая ведёт общение с пользователем через интерфейсную часть и выполняет запуск реализации алгоритма медицинской диагностики на МВС с последующей трансляцией её результатов к клиенту-пользователю.

Поскольку алгоритм диагностики для быстрой постановки диагноза должен иметь быстрый доступ к знаниям и данным с как можно меньшим количеством промежуточных трансляционных узлов/прослоек, то до начала диагностики на МВС отправляется наполнение базы знаний о заболеваниях, для чего в подсистему оптимизации этой базы знаний (публикатор) встроены выполняющий это код. При начале диагностики из ИН банка на МВС транслируется только история болезни пациента.

По окончании диагностики все полученные результаты транслируются с МВС сервер-частью решателя в интерфейсную клиент-часть.

В **пятой главе** приводятся данные экспериментального исследования свойств алгоритма медицинской диагностики при различных объёмах входных данных и знаний и различной конфигурации аппаратной платформы, а также исследование производительности прототипа банка знаний на реальных данных.

Полное исследование временной сложности алгоритма медицинской диагностики делится на две основные стадии: исследование зависимости времени работы алгоритма от объёма знаний предметной области (варьируется количество заболеваний и количество признаков) и от объёма данных (варьируется количество наблюдений значений признаков и количество происходящих событий).

При выполнении каждого этапа исследования варьируются три характеристики:

- основная варьируемая характеристика этапа исследования (см. выше),
- использование оптимизации по данным (сокращение перебора гипотез),
- использование параллелизма – варьирование количества запускаемых параллельно копий программы (одна – работает только сервер-процесс, две – работают сервер-процесс и один клиент-процесс, три и более – работают сервер-процесс и не менее двух клиент-процессов).

Исследование производительности алгоритма выполнялось на МВС Aleph, расположенной в ИАПУ ДВО РАН. Данная МВС работает под управлением ОС Linux, состоит из 4 узлов с псевдо-2-ядерными (HT) процессорами Intel Pentium4 3GHz и общей зеркалируемой файловой системой.

На рис. 2.а и 2.б представлены результаты измерения времени работы алгоритма при проведении исследования на первом этапе первой стадии – исследование зависимости времени работы алгоритма от количества заболеваний. При этом выбирается количество заболеваний, равное 0 (только знания о здоровом пациенте), 10, 20, 30 и 40. Помимо этого база знаний включает информацию о 12 анатомо-физиологических особенностях (по 3 значений у каждой), 100 признаках (по 6 значений у каждого), 10 событиях (по 6 значений у каждого). Наблюдения пациента включают значения всех особенностей, 8 событий и 5 моментов наблюдения каждого признака из 100. Данные пациента сгенерированы таким образом, чтобы одна из гипотез о диагнозе подтвердилась.

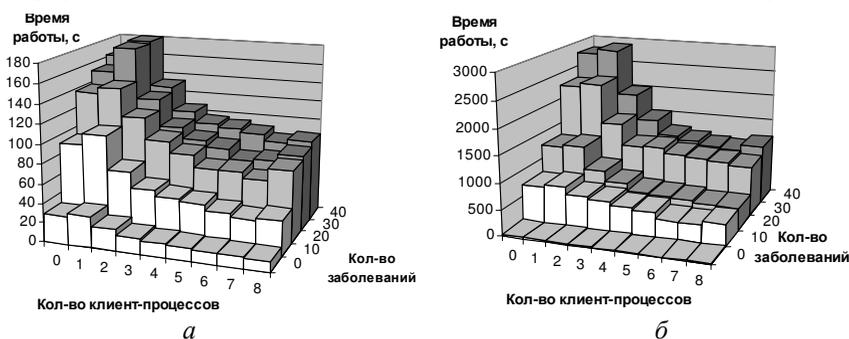


Рис. 2. Время работы алгоритма диагностики

при варьировании количества заболеваний в базе знаний

(а – с использованием оптимизации; б – без использования оптимизации)

За счёт оптимизации в случаях, когда количество заболеваний больше 0 (10, 20, 30 и 40), число исследуемых гипотез о заболевании сократилось до 5, 8, 12 и 13 соответственно (что составляет 2, 2,5, 2,5 и 3 раза соответственно).

Полученные значения говорят о том, что использование оптимизации по данным существенно сократило время постановки диагнозов (в среднем в 1-2 десятка раз). Данные, полученные при варьировании числа заболеваний в базе знаний, говорят о том, что при их увеличении временная сложность алгоритма растёт менее, чем линейно (вывод сделан на основе данных, полученных при использовании оптимизации и наиболее оптимальной загрузке МВС – при 7 клиент-процессах и сервер-процессе).

Для оценки практической применимости разработанного специализированного банка знаний в рамках платформы МБкЗ было сформированное его наполнение в области офтальмологии. Были созданы база знаний о наблюдениях и база знаний о заболеваниях с описанием 13 заболеваний и нормальной физиологии. В базу историй болезни банка были занесены 24 истории болезни реальных больных. В результате отладки базы знаний о заболеваниях система

во всех случаях выдала более конкретные диагнозы по сравнению с диагнозами, поставленными врачами. В случае недостаточного количества данных о пациенте помимо правильного диагноза выдавались и несколько альтернативных.

Процесс диагностики любой истории болезни занимает не более 30 сек, из которых 10-15 сек тратится на передачу истории болезни на МВС, около 5 секунд уходит на саму диагностику и 5-10 сек уходит на получение результатов диагностики на ПК пользователя. Формирование удобочитаемого объяснения полученных результатов занимает также порядка 30 сек.

Исследование производилось с использованием компьютера с ОС MS Windows XP, процессором Intel Pentium-4 HT 3ГГц, 2ГБ оперативной памяти.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

1. Разработана математическая модель онтологии медицинской диагностики острых заболеваний. Модель учитывает взаимодействие причинно-следственных связей различных типов. Онтология близка к реальным представлениям медицины в нашей стране и описывает сочетанную и осложненную патологии, динамику патологических процессов во времени, а также воздействие лечебных мероприятий и других событий на проявления заболеваний. Модель онтологии включает определения терминов модели знаний и модели действительности, а также систему соотношений, состоящую из ограничений целостности модели знаний и модели действительности и соотношений между ними.
2. Даны постановки прямой и обратной задач медицинской диагностики в терминах модели онтологии, а также постановка частной практической полезной обратной задачи, для которой разработан метод решения, допускающий оптимизацию по данным и естественное распараллеливание. Для алгоритма её решения доказаны теоремы об остановке, корректности и о функциональной эквивалентности исходной и модифицированной параллельной версий алгоритма.
3. Разработана концепция комплекса программ для решения задачи медицинской диагностики, названного банком знаний по медицинской диагностике. Выполнена его реализация в виде специализированного банка знаний в рамках платформы Многоцелевого банка знаний и многопроцессорной вычислительной системы, применяемой в качестве сервера приложений. Программные компоненты банка доступны через сеть Internet.
4. Проведено экспериментальное исследование вычислительной сложности разработанного алгоритма решения задачи медицинской диагностики, показавшее, что в широком диапазоне объёмов модельных баз знаний и историй болезни время его работы на одном процессоре не превышает нескольких минут, а использование многопроцессорной вычислительной системы и оптимизации сокращает время работы в несколько раз. Кроме того, исследование применимости реализованного комплекса программ по медицинской диагностике на реальных историях болезни (в области офтальмологии) подтвердило его практическую полезность.

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Клещёв А.С., Москаленко Ф.М., Черняховская М.Ю. Модель онтологии предметной области «Медицинская диагностика». Часть 1. Неформальное описание и определение базовых терминов. // **НТИ. Серия 2.** – 2005. – №12. – С.1-7.
2. Клещёв А.С., Москаленко Ф.М., Черняховская М.Ю. Модель онтологии предметной области «Медицинская диагностика». Часть 2. Формальное описание причинно-следственных связей, причин значений признаков и причин заболеваний. // **НТИ. Серия 2.** – 2006. – №2. – С.19-30.
3. Москаленко Ф.М. Задача медицинской диагностики и алгоритм её решения, допускающий распараллеливание. // **Информатика и системы управления.** – 2005. – №2(10). – С.52-63.
4. Москаленко Ф.М. Параллельный оптимизированный алгоритм медицинской диагностики. // **Информатика и системы управления.** – 2006. – №1(11). – С.87-98.
5. Москаленко Ф.М. Экспериментальное исследование временной сложности параллельного алгоритма диагностики, основанного на реальной онтологии медицины. // **Информатика и системы управления.** – 2006. – №2(12). – С.42-53.
6. Москаленко Ф.М. Проект компьютерного банка знаний по медицинской диагностике. // **Информатика и системы управления.** – 2007. – №2(14). – С.55-66.
7. Москаленко Ф.М. Компьютерный банк знаний по медицинской диагностике (в области офтальмологии). // Открытый дальневосточный конкурс программных средств студентов, аспирантов и молодых специалистов «Программист 2009». (Сборник докладов). – Владивосток. – 2009. – С.15-19.
8. Москаленко Ф.М., Черняховская М.Ю. База наблюдений в области офтальмологии для банка медицинских знаний. // **Информатика и системы управления.** – 2009. – №2(20). – С.40-49.
9. Москаленко Ф.М., Черняховская М.Ю. Формирование баз знаний о заболеваниях на основе онтологии медицины. // **Информатика и системы управления.** Материалы III научной конференции «Системный анализ в медицине». – Благовещенск. – 2009. – №4(22). – С.200-202.
10. Moskalenko Ph. M., Chernyakhovskaya M. Yu. An implementation of the computer bank of medical diagnostics for ophthalmology. // First Russia and Pacific Conference on Computer Technology and Applications (RPC-2010). – Vladivostok, Russia. – 2010. – pp.340-343.
11. Москаленко Ф.М., Черняховская М.Ю. База заболеваний «конъюнктивит» для компьютерного банка знаний. // **Информатика и системы управления.** – 2010. – №4(26). – С.94-103.

Личный вклад автора. Все результаты, составляющие основное содержание диссертационной работы, получены автором самостоятельно. В работах [1,2] автором разработана и формализована модель онтологии медицинской диагностики. В работах [8,9,11] автор формализовал знания в области офтальмологии. В работе [10] автор дал сведения о реализации программной системы и описал результаты исследования её производительности.

Москаленко Филипп Михайлович

**Методы решения задачи медицинской диагностики
на основе математической модели предметной области**

Автореферат

Подписано к печати 21.12.10
Формат 60×84/16

Усл.печ.л. 1.0
Тираж 100

Уч.-изд.л. 0.8
Заказ 40

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