



А.Н. ВИНОГРАДОВ,

к.ф.-м.н., заместитель руководителя Исследовательского центра искусственного интеллекта Института программных систем им. А.К. Айламазяна РАН, г. Переславль-Залесский, Россия, andrew@andrew.botik.ru

Я.И. ГУЛИЕВ,

к.т.н., руководитель Исследовательского центра медицинской информатики Института программных систем им. А.К. Айламазяна РАН, г. Переславль-Залесский, Россия, viit@yag.botik.ru

Е.П. КУРШЕВ,

к.т.н., руководитель Исследовательского центра искусственного интеллекта Института программных систем им. А.К. Айламазяна РАН, г. Переславль-Залесский, Россия, epk@epk.botik.ru

В.Л. МАЛЫХ,

к.т.н., зав. лабораторией Исследовательского центра медицинской информатики Института программных систем им. А.К. Айламазяна РАН, г. Переславль-Залесский, Россия, mvl@interin.ru

ПЕРСПЕКТИВНЫЕ НАПРАВЛЕНИЯ ИССЛЕДОВАНИЙ В ОБЛАСТИ КЛИНИЧЕСКОГО МОДЕЛИРОВАНИЯ, УПРАВЛЕНИЯ И ПРИНЯТИЯ РЕШЕНИЙ

УДК 61:007, 519.711.3

Виноградов А.Н., Гулиев Я.И., Куршев Е.П., Малых В.Л. *Перспективные направления исследований в области клинического моделирования, управления и принятия решений* (Институт программных систем им. А.К. Айламазяна РАН, г. Переславль-Залесский, Россия)

Аннотация. Статья посвящена перспективным исследованиям в области управления и поддержки принятия решений в клиническом процессе. Сформулированы основные научные проблемы, связанные с моделированием лечебно-диагностического процесса: а) проблема выбора подхода; б) проблема унификации модели; в) проблема полноты модели; г) проблема больших данных; д) проблема интеллектуального извлечения данных из профессиональных медицинских текстов на русском языке. Намечены пути решения указанных проблем. Обсуждается практическое значение данных исследований. Предлагается план масштабного проекта по практической реализации нового поколения информационных систем поддержки принятия решений и управления в медицине. Статья будет полезна архитекторам и разработчикам современных МИС.

Ключевые слова: медицинские информационные системы, системы автоматической обработки текстов, модель лечебно-диагностического процесса, генерализация медицинских данных, системы принятия решений в медицине, онтология, анализ текстов на естественном языке, структурирование текстовой информации, извлечение информации, классификация текстов, клинические записи, формализация медицинских сведений.

UDC 61:007, 519.711.3

Vinogradov A.N., Guliev Y.I., Kurshev E.P., Malykh V.L. *Perspective researches in the field of clinical modeling, control and decision making* (Ailamazyan Program Systems Institute of RAS, Pereslavl-Zalesky, Russia)

Abstract. Article is devoted to perspective research in the field of management and decision-making in clinical process. The main scientific problems connected with the modeling of diagnostic and treatment processes are formulated: a) problem of approach choice; b) problem of model unification; c) problem of model completeness; d) problem of big data; e) problem of information extraction from professional medical texts in Russian. Solutions of specified problems are outlined. Practical value of this research is discussed. Plan of large-scale project of implementation of new generation of decision-making support and management information systems in medicine is offered. The article will be useful to architects and developers of modern MIS.

Keywords: medical information system, automatic text processing systems, model of the diagnostic and treatment process, generalization of medical data, decision-making system in medicine, ontology, natural language analysis, textual information structuring, information extraction, text classification, clinical records, formalization of clinical data



Введение

Согласно мнениям экспертов, отраженных в аналитическом обзоре CNews «ИТ в здравоохранении 2014» [1], информационные технологии (ИТ) в медицине сами по себе не формируют новые способы лечения, но зато резко повышают эффективность как оказания медицинской помощи, так и функционирования объектов медицины. Будущее медицины за такими мощными средствами, как прогнозная аналитика и data mining (обнаружение новых знаний в данных), которые позволяют выявлять в накопленных данных определенные важные закономерности для прогнозирования возникающих тенденций. Традиционно прогнозная аналитика вместе с экспертными знаниями применяются в диагностике и при лечении многих болезней, обработка информации о низкой или высокой опасности возникновения определенного заболевания или состояния в группе пациентов с помощью средств интеллектуального анализа или прогнозной аналитики помогает врачам создавать протоколы лечения с учетом особенностей разных групп. На передний край ИТ в медицине будущего выходят проблемы поддержки принятия врачебных решений, а также предоставление населению интеллектуальных средств контроля своего здоровья (прогнозирование критических ситуаций, выработка рекомендаций по их предупреждению, контроль за принимаемыми лекарственными средствами).

Долгое время попытки использования математических методов в медицине для решения задач постановки диагноза, прогнозирования развития и исхода клинического процесса, предупреждения критических ситуаций, принятия решений и управления клиническим процессом приводили лишь к отдельным частным результатам, имеющим практическое значение в ограниченном контексте частной нозологии. Этот подход к алгоритмизации клинической работы врачей отражен в работах известного математика

И.М. Гельфанда и его школы [2]. Сложности, связанные с построением частной модели клинического процесса для отдельной нозологии, сложности в идентификации параметров таких моделей не способствовали широкому практическому внедрению моделей клинического процесса в современные медицинские информационные системы (МИС).

В настоящее время мы являемся свидетелями расширения подходов к решению проблемы управления и поддержки принятия решений в клиническом процессе. Наряду с модельным подходом, предполагающим детальную разработку частных нозологических моделей, предлагается также начать широко использовать прецедентный подход. Основное требование прецедентного подхода — это наличие достаточного числа релевантных прецедентов.

Важно отметить прецедентный характер управления в медицине [3–5]. Медицине свойственен консерватизм в принятии решений: активные действия (управление) зачастую выбираются на основе уже известных прецедентов, доказавших свою эффективность в статистическом смысле (доказательная медицина), при этом также учитывается история процесса. Схожесть и повторяемость ситуаций приводят к схожести и повторяемости управления в этих ситуациях. Прецеденты становятся носителями знаний о том, как следует поступить, как поступали ранее в данной ситуации. Прецедентный подход известен давно, но условия для практического применения этого подхода в медицинской информатике сложились только сейчас. Появились информационные технологии, позволяющие работать с большими данными, и появились сами данные, накопленные в медицинских информационных системах. Открылась возможность практического применения прецедентного подхода для решения проблемы управления и поддержки принятия врачебных решений в клиническом процессе.

Отметим особенность постановки задачи управления для прецедентного подхода [3–4].





В классической постановке задачи управления мы должны были бы указать возможные значения управления и связать между собой динамику состояния объекта управления $x(t)$ и управление $u(t)$, например, в виде следующего уравнения $\dot{x} = f(x, u)$, $x \in X$, $u \in U$. Задача управления по прецедентам имеет особенность, отличающую ее от классической постановки задачи управления. Множество U можно считать заданным, но функция $f(x, u)$ нам «известна» только лишь прецедентно, то есть известна только для наблюдавшихся ранее ситуаций (x, u) . На первый план для построения управления выходит задача поиска «близкого» релевантного прецедента.

Итак, проблемы принятия решений в лечебно-диагностических процессах долгое время рассматривались как проблемы исключительно самой медицины, и информационные технологии имели к их решению лишь очень далекое косвенное отношение. В настоящее время ситуация стремительно меняется. Произошло осознание того, что современные информационные технологии в медицине позволяют накапливать и хранить огромное множество медицинских фактов. Эти факты заключают в себе медицинские знания, как актуальные, так и в ретроспективе. Основная проблема — это наше «неумение» извлекать эти знания и использовать их для принятия решений.

В работе приводятся первые результаты применения прецедентного подхода к моделированию клинического процесса и управлению им на основе реальных клинических данных.

Проблема выбора подхода

Существуют два основных прикладных подхода к построению систем поддержки принятия решений: принятие решений по прецедентам и по моделям [6]. Оба подхода считаются равноправными, и каждый имеет свои достоинства и недостатки. При модельном подходе, «как и при всяком обобщении, теряются некоторые особенности поведения

системы в каждой конкретной точке пространства решений. При опоре же на прецедент, как показывает опыт, удается учесть эти локальные особенности, что часто позволяет получать более точные решения» [6]. Исторически именно прецедентный подход был положен в основу медицины, начиная еще с древнегреческих времен. Эффективность прецедентного подхода к управлению и принятию решений в медицине, эффективность многоаспектного использования прецедентов в медицинских информационных системах всячески подчеркивается и обосновывается в работах [3–5, 7]. По нашему глубокому убеждению, именно прецедентный подход позволит создать в ближайшей перспективе эффективные и точные системы поддержки принятия врачебных решений. Не замыкаясь в частностях построения моделей отдельных нозологий, прецедентный подход предлагает общий универсальный путь решения проблемы поддержки принятия врачебных решений. Основное требование прецедентного подхода — это наличие достаточного числа релевантных прецедентов. Из этого требования вытекает необходимость формализации и накопления в информационных хранилищах медицинских фактов как в национальном масштабе, так в перспективе и в общемировом. Здесь мы неизбежно должны будем столкнуться с проблемой Больших Данных (Big Data).

Проблема Больших Данных

Направление информатики, посвященное Большим Данным, несмотря на молодость, считается одним из самых высокоприоритетных в области ИТ. У больших данных в медицинской информатике есть своя специфика. В отличие от данных БАК (большого андронного коллайдера), медицинских данных (под «элементом» данных здесь понимается описанная во всей полноте одна реализация лечебно-диагностического процесса) будет меньше, порядка десяти в девятой или в десятой степени, но при этом все эти данные



будут значимы. Размерность данных (число различных характеристик, описывающих состояние человека) весьма велика, порядка десяти в третьей или в четвертой степени. Это очень сильно отличает эти данные от физических больших данных, получаемых с Большого адронного коллайдера или ожидаемых от телескопа Large Synoptic Survey Telescope, наиболее часто приводимых в качестве примера в работах по Большим Данным. Если за Большими Данными, получаемыми с физических приборов, стоят соответствующие разработанные теории и модели (стандартная теория), то медицина не может похвастаться подобными основами. В вычислительном плане, благодаря суперкомпьютерам и технологиям распределенных вычислений, ИТ научились справляться с большими данными. Но для нашей предметной области на передний план выходит проблема концептуализации больших медицинских данных. Никакой общепринятой формализованной стандартной модели лечебно-диагностического процесса в настоящее время не существует [2–5, 8–9]. У медицинских фактов огромное множество источников, «говорящих» на своих профессиональных языках и диалектах. Очевидно, что для решения проблемы больших данных в медицинской информатике требуется искать свои предметно-ориентированные подходы. В первую очередь необходимо построить некоторую *стандартную* модель ЛДП.

Проблема унификации модели ЛДП

С точки зрения кибернетики, дискретная модель динамического процесса может иметь определенные преимущества перед непрерывной моделью [10]. Математика дает нам достаточно общие и широко известные формализации дискретного динамического процесса [11]. В самом общем случае дискретный динамический процесс — это конечная последовательность состояний. Переходы между состояниями осуществляются в некото-

рые, возможно случайные, моменты времени. Переходы определяются как внутренней динамикой системы, так и внешними факторами, в том числе управлением со стороны других систем. Само управление может зависеть не только от текущего состояния системы, но и от более ранних состояний, от предыстории динамического процесса. Еще одна важная характеристика динамического процесса — является ли он детерминированным или стохастическим. Если процесс детерминирован, то две его реализации, совпадающие между собой по текущим состояниям и по всей своей предыстории, при одном и том же управлении должны будут перейти в одно и то же состояние. В стохастическом процессе переходы из состояния в состояние будут происходить случайно с некоторыми вероятностями. Существует общий кибернетический взгляд на эту проблему, сформулированный Р. Эшби. Он заключается в том, что все *сложные* системы и протекающие в них процессы по необходимости должны иметь стохастическую природу. Одно из возможных объяснений природы стохастичности процесса — это наличие скрытых ненаблюдаемых характеристик состояния, влияющих на динамику процесса (точка зрения А. Эйнштейна на случайность в физике). Очевидно, что ЛДП, отражающий динамику состояния человека, будет обладать всеми чертами сложного динамического процесса. ЛДП будет стохастическим, управление в нем будет зависеть от предыстории процесса, наблюдение за состоянием (скорее за отдельными характеристиками состояния) будет происходить в различные случайные моменты времени, на динамику процесса будут влиять скрытые ненаблюдаемые параметры. С неизбежностью мы должны будем упростить модель процесса, предлагаемую в качестве стандартной модели ЛДП, не потеряв при этом основные черты процесса. В работах [3–4] была предложена модель ЛДП, которая претендует стать основой стандартной унифици-





рованной модели ЛДП. При разработке модели был принят ряд важных концептуальных решений.

Во-первых, предполагается равномерное темпирование процесса, его приведение к дискретному процессу с постоянным шагом по времени. При этом мы сталкиваемся с проблемой приведения значений характеристик состояния, наблюдаемых с отличным от равномерного темпом, к некоторым «темперированным» значениям. При этом приведенная к равномерному темпу характеристика состояния может даже по типу отличаться от своих первичных наблюдаемых значений. Например, для темпированного процесса с шагом в 24 часа (соответствует нахождению пациента в стационаре) при нескольких измерениях температуры тела за день в качестве приведенного значения температуры вместо малоосмысленной «средней температуры за день» мы можем дать интерпретацию суточных колебаний температуры: «температура повышается», «больного лихорадит» и т.п. Равномерное темпирование процессов крайне важно, так как в результате мы получаем возможность соотнесения состояний различных реализаций процессов между собой для заданного шага процесса, другими словами, с учетом времени, прошедшего от начала реализации процесса. Приведение значений характеристик состояния пациента к одному темпу позволяет нам как бы «собрать» состояние из множества характеристик, наблюдаемых с различными темпами в произвольные моменты времени. А понятие состояния для нас является базовым, так как в дальнейшем мы будем решать задачу поиска близкого релевантного состояния — прецедента.

Во-вторых, для уменьшения эффекта памяти процесса, предполагающего, что управление зависит как от текущего состояния, так и от всей предыстории процесса, предполагается выполнять операцию «интегрирования» управления и включать интегральные характе-

ристики управления в расширенное состояние. Цитируем [4]: «Содержательно управление — это выбор врачом на различных шагах процесса различных лечебно-диагностических воздействий. С каждым лечебно-диагностическим воздействием (элементом) можно сопоставить некоторую интегральную характеристику применения этого элемента в ЛДП. Например, для лекарственного средства такой характеристикой будет суммарная принятая пациентом доза, для лучевой терапии — суммарная доза облучения, интегральной характеристикой часто может быть кратность применения данного элемента, например, число проведенных электрокардиографических исследований. Интегральные характеристики управления включаются в состояние объекта управления и «ослабляют» эффект памяти, мы «приближаемся» к процессу, в котором управление становится функцией состояния, расширенного интегральными характеристиками управления».

В-третьих, учитывая большую размерность пространства состояний, даже взяв достаточно большой (от десяти в четверной степени и выше) ансамбль реализаций ЛДП для одной нозологии, формально мы не «увидим» в этом ансамбле стохастического поведения. Так как состояния для различных реализаций, взятые на одном временном шаге процесса, пусть и незначительно, но будут отличаться друг от друга. Для появления в модели стохастичности предлагается специальная процедура генерализации характеристик состояния. На множестве значений каждой характеристики в рамках данной нозологии вводятся отношения эквивалентности, которые разбивают множество значений характеристики на классы эквивалентности. В результате применения генерализации «стираются» незначительные формальные различия между состояниями, реализации ЛДП из представительного ансамбля начинают разделять между собой различные генерализованные состояния, и ожидается проявление стохастического характера процесса.



Предложенная модель ЛДП не является окончательной, продолжается ее развитие и уточнение. В частности, предполагается явное включение в модель различных интерпретаций состояния. Модель ЛДП должна позволять осуществлять логический вывод на основе многозначных логико-математических моделей с возможностью множественных интерпретаций. Речь идет о принятии различных решений относительно данного состояния, например, о постановке диагноза, или об отнесении состояния к одной или нескольким критическим ситуациям. Результаты интерпретации состояния должны быть включены в предлагаемую модель ЛДП, но не в виде независимых характеристик состояния, а в виде функций состояния $f_i(x)$. Машинная интерпретация состояния должна выполняться на основе многозначных логик (логика Лукасевича) или бесконечнозначных логик. Сами процессы принятия решения человеком и накопления достаточной для принятия решения информации будут отражаться в реализациях — прецедентах ЛДП. В частности, по прецедентам можно будет изучать структуры диагностической информации, достаточной для принятия различных диагностических решений.

Проблема полноты данных

Медицина устроена так, что для нее важны любые, как часто встречающиеся « типовые » клинические случаи, так и редкие « особые » клинические прецеденты. Пример из [1]: « детские онкологические заболевания — отрасль очень узкая, а если брать еще уже — заболевания конкретных органов — так информации вообще практически нет, особенно в регионах. У любого районного онколога должен быть доступ к данным о том, что такой-то прецедент в практике здравоохранения уже имел место и при этом был назначен такой-то способ лечения ». Информация такого рода поможет диагностировать различные заболевания на ранних стадиях и выбирать оптимальные варианты лечения быстро.

А время в медицине является ключевым фактором.

Очевидно, что полнота данных нами рассматривается в двух аспектах. С одной стороны, эффективность прецедентного подхода будет определяться мощностью накопленных в банке клинических данных реализаций ЛДП. Как мы уже отмечали выше, желательно иметь по каждой отдельной нозологии не менее десяти в третьей — десяти в четвертой степени реализаций. С другой стороны, нас также заботит полнота описания отдельной реализации ЛДП, полнота описания состояния. В современных медицинских информационных системах лечебно-диагностический процесс фиксируется как развернутая во времени последовательность событий-управлений (к ним относятся различные лечебно-диагностические назначения) и событий-наблюдений (они характеризуют состояние пациента). События-управления достаточно хорошо формализованы, ведется их статистический и экономический учет: формируются реестры оказанных услуг, выставляются счета, производятся планирование и диспетчеризация. С событиями-наблюдениями ситуация значительно хуже. Если результаты лабораторных исследований помещаются в базу данных медицинской информационной системы в относительно структурированном виде, то результаты различных инструментальных диагностических исследований и собственно врачебные наблюдения, как правило, присутствуют в системе в виде свободных неформализованных текстов. Отсутствие возможности автоматического анализа свободных медицинских текстов приводит к элиминированию содержащихся в них фактов из математических моделей лечебно-диагностического процесса, снижает возможности поиска таких фактов в массивах клинических данных. В работе [12] мы уже отмечали, что возможности структурированного ввода данных в МИС ограничены, неструктурированные свободные медицинские тексты будут всегда





присутствовать в данных МИС. Необходимо решать проблему интеллектуального извлечения данных из текстов на естественном профессиональном языке.

Проблема интеллектуального извлечения данных

Проблема состоит в том, чтобы превратить неструктурированные текстовые записи в структуры данных, пригодные для использования в разнообразных целях: различные виды поиска (например, поиск группы пациентов, отвечающих заданным критериям; получение сводной выборки сведений о пациенте, содержащихся в разных электронных документах), поддержка принятия решений, управление лечебно-диагностическим процессом, управление ресурсами, вопросно-ответные системы. Объединение результатов анализа разнородных медицинских текстов, включающих клинические записи, научные публикации, нормативную документацию, описания лекарственных препаратов, с имеющимися структурированными данными расширяет возможности использования таких методов аналитической обработки данных, как data mining. Применительно к медицине это означает открытие ранее неизвестных корреляций между симптомами и заболеваниями, обнаружение неописанных побочных эффектов от применения лекарственных препаратов и т.п.

В последние 30 лет в мире отмечается активный рост интереса к анализу естественного языка в медицине и смежных областях [13–17]. Начиная с 80–90-х годов, системы анализа медицинских текстов получили возможность опираться на общедоступные предметные ресурсы. Самым масштабным унифицированным терминологическим и онтологическим ресурсом для систем анализа медицинских текстов можно считать UMLS (Universal Medicine Language System). Это интегрированный ресурс, включающий в себя три компонента: метатезаурус, семантическую сеть (множество концептов и связей между

ними) и компонент Specialist SPECIALIST Lexicon and Lexical Tools (лингвистический словарь терминов и программное средство для распознавания терминов в тексте). Метатезаурус UMLS объединяет более 150 ресурсов: онтологий, тезаурусов, терминологий, наборов кодов и классификаций, в том числе ICD-9 (Международную классификацию болезней), SNOMED-CT и MeSH. SNOMED-CT — систематизированную медицинскую номенклатуру (онтология и многоязычный тезаурус для медицины и здравоохранения). Согласно данным IHTSDO (The International Health Terminology Standards Development Organisation), медицинская номенклатура SNOMED-CT в настоящее время насчитывает более 300 тыс. концептов (<http://www.ihtsdo.org/snomed-ct/snomed-ct0/snomed-ct-components/>). Тезаурус MeSH разработан Национальной библиотекой медицины США (NLM). Дескрипторы MeSH используются не только в каталогах самой NLM, но и для индексирования документов в крупнейшей библиографической базе данных по биомедицине MEDLINE (входит в ресурс PubMed, поддерживаемый NLM). Унифицированного русскоязычного ресурса, аналогичного UMLS, не существует.

В настоящее время основным подходом, применяемым при анализе свободного текста (narrative) в электронных медицинских картах, является извлечение информации (Information Extraction). Извлечение информации — это обнаружение в тексте и представление в структурированном виде информации некоторого заранее заданного типа. Анализу подвергаются анамнезы, выписные эпикризы, отчеты об обследовании (рентгенологических, ЭКГ) и др. Довольно успешно (на пригодном для практического использования уровне) извлекается демографическая информация, упоминания процедур, препаратов, симптомов, признаков, заболеваний, диагнозов. Задача обезличивания (деперсонификации, деидентификации) данных электронной медицинской карты для получения коллекций



текстов также относится к задачам извлечения информации, поскольку требует обнаружения и удаления из текста определенных категорий данных (имена, адреса, номера телефонов, различные числовые идентификаторы, даты и т.п.). Несмотря на то, что извлечение информации относится к так называемым поверхностным (узконаправленным) подходам к автоматической обработке текста, системы обработки клинических записей стоят перед необходимостью практически пригодного решения сложнейших проблем в области компьютерного анализа естественного языка. К ним относятся проблема языковой неоднозначности на всех уровнях анализа [18, 19], задача отождествления (установления кореферентности) текстовых упоминаний [20], задача моделирования временных (темпоральных) аспектов информации [21]. Технология извлечения информации особенно эффективна при решении задач, требующих обработки большого объема данных — таких как отслеживание нежелательных явлений на фоне лечения, побочных эффектов лекарственных препаратов, синдромный надзор и пр.

Актуализация и концептуализация знаний, формирование онтологии предметной области

В более ранних наших работах мы уже отмечали целесообразность и необходимость встраивания механизма концептуализации предметной области в архитектуру МИС [22]. Цитируем: «Основная концептуальная идея архитектуры ... введение процесса концептуализации предметной области непосредственно в саму информационную систему». Процесс концептуализации в указанной работе рассматривался только с точки зрения построения моделей клинических документов МИС. Справедливо отмечалось, что клинические документы являются важнейшими источниками знаний о предметной области. Теперь эту точку зрения можно расширить. К источникам знаний — моделям клинических доку-

ментов следует добавить знания, заключенные в унифицированной модели лечебно-диагностического процесса в целом [3–5]. Здесь мы неизбежно сталкиваемся с общесистемными проблемами. Речь идет о необходимости усвоения и добавления в систему новых знаний, о построении в системе непрерывного процесса концептуализации и формализации нового знания. Оказывается, что можно предложить достаточно общий подход для решения этой общесистемной проблемы. Этот подход, получивший название прецедентного подхода, изложен в работе [7]. Появление нового прецедента в МИС равносильно появлению новых знаний в системе, за которым следует возникновение соответствующих событий, например, уведомления о необходимости контроля нового прецедента — необходимости оценки корректности прецедента, необходимости концептуализации и ввода в систему новых знаний для корректных прецедентов. Появление новых знаний в системе тут же открывает возможности их немедленного функционального использования. Для клинических документов прецеденты могут заключаться в появлении новых документов, в появлении в структуре документа ранее не формализованных понятий и концептов. В лечебно-диагностическом процессе новые прецеденты — это ранее не наблюдавшиеся события (факты).

Численное моделирование ЛДП по реальным клиническим данным

Далее приводим результаты, впервые представленные в [4], дополненные еще одной нозологией. Силами сотрудников ИПС им. А.К. Айламазяна РАН выполнялись вычислительные эксперименты с предложенной математической моделью ЛДП. Целью экспериментов была проверка концептуальных идей, заложенных в модель. Было высказано общее предположение о том, что с ростом объема моделируемого ансамбля процессов





Таблица 1

Ансамбли моделируемых лечебно-диагностических процессов

№	Код нозологии	Название нозологии	Число реализаций процессов	Число состояний/ Число обобщенных состояний	Сжатие числа состояний после их обобщения	Мощность словаря до нормализации/ после нормализации
1	J13	Пневмония, вызванная Streptococcus pneumoniae	166	2938/2921	<1%	828/128
2	H26.2	Осложненная катаракта	1255	5778/2308	60%	328/249
3	I20.8	Другие формы стенокардии	3069	48 909/48 513	<3%	871/99
4	I10	Эссенциальная [первичная] гипертензия	8734	98 389/82 542	16%	3223/1278

будет наблюдаться все большее «сжатие» числа модельных состояний по отношению к числу исходных состояний, все большее повторение состояний в реализациях процессов. Сложности в моделировании добавляла необходимость нормализации словарей характеристик, выделяемых из реальных клинических данных. Качественно выполнить работу по нормализации словарей можно было только с участием врачей-экспертов.

Были рассмотрены три ансамбля процессов, см. таблицу 1.

В таблице 1 представлены первые результаты численного моделирования лечебно-диагностических процессов. Указан код нозологии по международному классификатору ICD 10-го пересмотра, указано название нозологии. По каждой из представленных нозологий отбирались законченные клинические случаи госпитализации, в которых основной диагноз кодировался данной нозологией. Указано число процессов (законченных случаев), отобранных по данной нозологии. Временной шаг в дискретных модельных процессах был принят равным 24 часам. Каждый из рассматриваемых процессов длился целое число дней. Состояние пациента фиксировалось один раз в сутки. Каждое из состояний описывалось некоторым набором в общем слу-

чае размерных характеристик с определенными значениями. Совокупность всех зафиксированных для данного ансамбля процессов характеристик формировала словарь характеристик для данной нозологии. Мощности трех словарей характеристик для трех нозологий представлены в таблице. В таблице 1 представлена часть словаря характеристик для нозологии J13. Проводилась нормализация словарей характеристик. Нормализация сводилась к исключению из рассмотрения характеристик, которые не были связаны с основным заболеванием (данной нозологией) и не оказывали влияния на протекание основного заболевания. Также при нормализации решалась проблема синонимии, каждая характеристика получала некоторое стандартное наименование. Нормализацию словаря по нозологии J13 выполнил врач-эксперт. Нормализация словаря по нозологии H26.2 была выполнена силами авторов статьи. Для предварительной нормализации словаря для нозологии I20.8 был использован статистический подход, были исключены из словаря для данной нозологии все характеристики, которые имели рейтинг ниже 307, то есть встречались не более чем в 10% процессов из данного ансамбля. Безусловно, такой подход требует одобрения со стороны



врачей-экспертов, можно допустить такой подход на стадии предварительной грубой оценки возможностей предложенной модели. Мощности словарей характеристик после нормализации указаны в *таблице 1*. Отметим, что нормализация словарей характеристик влияет на описание состояний, но не оказывает никакого влияния на само число описываемых состояний. Для оценки потенциальной возможности сжатия описания ансамбля процессов после выполнения описанной в методологии моделирования процедуры построения обобщенных классов для характеристик использовался следующий подход. Считалось, что любое значение каждой характеристики отображается всего в один класс эквивалентности (предельное сжатие). Два состояния считались эквивалентными, если они имели один временной дискретный индекс шага процесса и имели одинаковые множества характеристик, описывающих эти два состояния. Число таких обобщенных состояний для каждой из нозологий представлено в таблице. Сжатие определялось как доля в процентах уменьшения числа состояний при переходе от начального описания процессов к обобщенному от начального числа состояний. Из общих соображений предполагалось, что сжатие будет незначительным при малом числе процессов в моделируемом ансамбле, в этом случае индивидуальные черты процесса будут отделять его от других не похожих на него. К этому случаю можно отнести нозологию J13, всего 166 процессов, менее 1% сжатия. По мере увеличения мощности моделируемого ансамбля процессов сжатие будет расти, для нозологии H26.2 с ансамблем из 1255 процессов было получено сжатие в 60%. Для нозологии I20.8 с ансамблем из 3069 процессов было получено незначительное сжатие, менее 3%. Это можно объяснить недостаточно качественной нормализацией словаря характеристик для данной нозологии, недостаточной мощностью ансамбля процессов, а также широтой нозологии.

Делать окончательные выводы по первым результатам численного моделирования рано. Исследования надо продолжить, увеличивая мощность ансамблей ЛДП, а также усиливая нормализацию и генерализацию первичных данных.

Заключение

В работе приводится аргументация в пользу смены подхода к решению проблемы управления и поддержки принятия решений в клиническом процессе. Предлагается решать указанную проблему с помощью прецедентного подхода. Приводятся ссылки на работы, посвященные применению прецедентного подхода для моделирования клинического процесса и управлению им на основе реальных клинических прецедентов. Работа поддержана грантом РФФИ 13-07-12012.

В заключение приведем план задач по реализации масштабного проекта в области поддержки принятия решений в клиническом процессе:

- 1.** Разработка математической модели, которая может служить стандартной теорией системы знаний и самого лечебно-диагностического процесса.
- 2.** Разработка методологии формализации и нормализации медицинских фактов с учетом быстрого изменения систем оценок и понятий, стандартизация языка описания фактов.
- 3.** Разработка масштабируемой архитектуры банка клинических данных от масштаба отдельной медицинской организации до национального масштаба.
- 4.** Разработка методов релевантного поиска клинических прецедентов.
- 5.** Разработка методов поддержания актуальности и целостности банка клинических данных в течение длительных промежутков времени.
- 6.** Реализация программного комплекса масштабируемого банка клинических данных.
- 7.** Разработка и реализация программных средств поддержки принятия решений на





основе прецедентного подхода, онтологической модели предметной области и унифицированной модели лечебно-диагностического процесса.

Выполнение этого плана позволит перейти к практической реализации нового поколения информационных систем поддержки принятия решений и управления в медицине.

ЛИТЕРАТУРА



1. Обзор: ИТ в здравоохранении 2014. CNews аналитика. <http://www.cnews.ru/reviews/new/publichealth2014/>.
2. Гельфанд И.М., Розенфельд Б.И., Шифрин М.А. Очерки о совместной работе математиков и врачей. — М.: Едиториал УРСС, 2005.
3. Малых В.Л., Гулиев Я.И. Управляемый стохастический прецедентный процесс с памятью как математическая модель лечебно-диагностического процесса//Информационные технологии и вычислительные системы. — 2014. — № 2. — С. 60–72.
4. Малых В.Л., Гулиев Я.И., Еремин А.В., Рудецкий С.В. Управление и принятие решений в лечебно-диагностическом процессе//В кн. XII Всероссийское совещание по проблемам управления ВСПУ-2014, Москва. Труды. [Электронное издание], 6518.pdf, с.6518–6528, <http://vspu2014.ipu.ru/node/8581>.
5. Малых В.Л., Гулиев Я.И. Моделирование лечебно-диагностического процесса в классе управляемых стохастических процессов с памятью//Врач и информационные технологии. — 2013. — № 2. — С. 6–15.
6. Загоруйко Н.Г. Прикладные методы анализа данных и знаний. — Новосибирск: ИМ им. С.Л. Соболева СО РАН, 1999.
7. Малых В.Л., Гулиев Я.И. Прецеденты в медицинских информационных системах// Программные продукты и системы. — 2009. — № 2(86). — С.19–27.
8. Назаренко Г.И., Осипов Г.С. Основы теории медицинских технологических процессов. Том 1. — М.: Физматлит, 2005.
9. Bernd Blobel. Clinical modeling-A critical analysis//International journal of medical informatics. — 2014. — № 83. — P. 57–69. [http://www.ijmijournal.com/article/S1386-5056\(13\)00201-3/pdf](http://www.ijmijournal.com/article/S1386-5056(13)00201-3/pdf).
10. Турчин В.Ф. Феномен науки. — М.: ЭТС, 2000.
11. Миронов А.М. Теория процессов. <http://intsys.msu.ru/staff/mironov/processes.pdf>.
12. Малых В.Л., Гулиев Я.И., Калинин А.Н., Колупаев А.В., Юрченко С.Г. Возможности применения речевого интерфейса и систем автоматической обработки текстов в МИС//Врач и информационные технологии. — 2014. — № 5.
13. Pai, 2014, Workshop on using natural language processing applications for enhancing clinical decision making: an executive summary, Journal of the American Medical Informatics Association: JAMIA.
14. Friedman, 2013, Natural language processing: State of the art and prospects for significant progress, a workshop sponsored by the National Library of Medicine, Journal of Biomedical Informatics; <http://www.ncbi.nlm.nih.gov/pubmed/23810857>.
15. Patrick, 2010, High accuracy information extraction of medication information from clinical notes: 2009 i2b2 medication extraction challenge, Journal of the American Medical Informatics Association : JAMIA; <http://jamia.bmj.com/content/17/5/524>.



- 16.** Meystre, 2008, Extracting Information from Textual Documents in the Electronic Health Record: A Review of Recent Research, IMIA Yearbook 2008: Access to Health Information <http://www.schattauer.de/en/magazine/subject-areas/journals-a-z/imia-yearbook/imia-yearbook-2008/issue/special/manuscript/9830/show.html>.
- 17.** Savova, 2010, Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications, Journal of the American Medical Informatics Association: JAMIA; <http://www.ncbi.nlm.nih.gov/pubmed/20819853>.
- 18.** Piantadosi S.T., Tily H., Gibson E. The communicative function of ambiguity in language//Cognition. — 2012. — № 122(3). — P. 280–291.
- 19.** Богуславский И.М., Иомдин Л.Л., Лазурский А.В., Митюшин Л.Г., Бердичевский А.С. Интерактивное разрешение внутренней и переводной неоднозначности в системе машинного перевода//В кн. Компьютерная лингвистика и интеллектуальные технологии (Диалог 2005). Труды конференции. Звенигород, 1–6 мая 2005 г. — М.: Наука, 2005. — С. 216–221.
- 20.** Bryan Rink, Kirk Roberts, Sanda M Harabagiu. A supervised framework for resolving coreference in clinical records//Journal of the American Medical Informatics Association
- 21.** Hanauer D.A., Ramakrishnan N. Modeling temporal relationships in large scale clinical associations//J Am Med Inform Assoc. — 2013. — Mar-Apr. — № 20(2). — P. 332–341.
- 22.** Бериков В.Б. Современные тенденции в кластерном анализе//[Электронный ресурс] Федеральный портал по научной и инновационной деятельности. Статьи — победители Всероссийского конкурсного отбора обзорно-аналитических статей по приоритетному направлению «Информационно-телекоммуникационные системы» (2008-07-09). <http://www.ict.edu.ru/ft/005638/62315e1-st02.pdf>.