

УДК (61:007)

УПРАВЛЕНИЕ И ПРИНЯТИЕ РЕШЕНИЙ В ЛЕЧЕБНО-ДИАГНОСТИЧЕСКОМ ПРОЦЕССЕ

В.Л. Малых*ИПС им. А.К. Айламазяна РАН*

Россия, 152021, Ярославская обл., Переславский район, с. Веськово, Петра I ул., 4а

E-mail: mvl@interin.ru**Я.И. Гулиев***ИПС им. А.К. Айламазяна РАН*

Россия, 152021, Ярославская обл., Переславский район, с. Веськово, Петра I ул., 4а

E-mail: vspu@yag.botik.ru**А.В. Еремин***ИПС им. А.К. Айламазяна РАН*

Россия, 152021, Ярославская обл., Переславский район, с. Веськово, Петра I ул., 4а

E-mail: artem_eremin@interin.ru**С.В. Рудецкий***ИПС им. А.К. Айламазяна РАН*

Россия, 152021, Ярославская обл., Переславский район, с. Веськово, Петра I ул., 4а

E-mail: rsv@interin.ru

Ключевые слова: лечебно-диагностический процесс, генерализация медицинских данных, математическая модель, стохастический управляемый процесс с памятью, Марковский процесс

Аннотация: В работе развивается ранее предложенная постановка задачи управления и принятия решений в лечебно-диагностическом процессе. Приводится математическая модель лечебно-диагностического процесса. Управление предлагается строить на основе известных клинических прецедентов. В пространстве состояний и процессов вводятся различные метрики для отбора прецедентов. Приводятся результаты численного моделирования на основе реальных клинических данных. Работа будет полезна специалистам из области медицинской информатики и кибернетики. Работа поддержана грантом РФФИ 13-07-12012.

1. Введение

Идеи о возможности моделирования лечебно-диагностического процесса (ЛДП) классом стохастических управляемых Марковских процессов были нами высказаны в работе [1]. Дальнейшее развитие этих идей привело к построению математической модели лечебно-диагностического процесса [2]. Для дальнейшего понимания работы нам необходимо будет сжато изложить основные идеи и привести математическую формулировку из указанных источников.

В ходе ведения лечебно-диагностического процесса врач преследует определенные цели, предпринимает активные лечебно-диагностические действия, влияющие на со-

стояние пациента. Активные действия, направленные на достижение цели, принято называть управлением. Быть может, в данном контексте использования, лучше подходит термин регулирование состояния пациента, регулирование здоровья, но мы остановимся на управлении. Сошлемся на Норберта Винера, который подчеркивал возможность применения теории управления к изучению и моделированию живых организмов. Подчеркнем «неклассический» характер рассматриваемой задачи. 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 $\dot{x} = f(x, u)$, $x \in X$, $u \in U$. Трудности подобной постановки задачи в медицине начнутся с первых же шагов. Начнем с понятия состояния объекта управления. Состояние пациента может описываться множеством различных симптомов и показателей. Наблюдение и определение характеристик состояния производится с различным темпом. Например, в стационаре лечащий врач раз в день совершает обход и осматривает пациента, Температуру у пациента могут измерить 2-3 раза в день. Рентгенологическое исследование при пневмонии проведут 2 раза за весь срок пребывания в стационаре. Результаты лабораторных исследований станут известны с некоторым, иногда весьма большим, запаздыванием. В реанимации темп измерения различных показателей достаточно высок. Для произвольного момента времени t довольно сложно дать «точный» ответ о состоянии $x(t)$. Обычные методы интерполяции и интегрирования быстро меняющейся переменной здесь плохо работают. Врача часто интересует не среднее значение характеристики за день, а ее динамика. Итак, различные характеристики состояния измеряются в различные моменты времени с различным темпом. Результаты измерений могут быть получены с запаздыванием. Перейдем теперь к рассмотрению особенностей динамики и управления и процитируем [2].

«Формально можно считать, что врач на каждом шаге ЛДП выбирает в качестве управления некоторое подмножество элементов ЛДП (диагностические исследования, лабораторные тесты, назначения лекарственных средств и т.п.), и выбор врача осуществляется на основании накопленных медициной знаний о том, как надо лечить данное заболевание, на основании собственного опыта врача. Не все подмножества из элементов ЛДП рассматривает врач в качестве кандидатов на управление в данной ситуации, но лишь те из них, которые применялись в схожих ситуациях в прошлом, которые доказали свою клиническую эффективность. Управление носит ярко выраженный прецедентный характер, а задача управления имеет особенность, отличающую ее от классической постановки задачи управления. Множество U можно считать заданным, но функция $f(x, u)$ нам «известна» только лишь прецедентно, т.е. известна только для наблюдавшихся ранее ситуаций (x, u) ».

Сложности построения математической модели динамики состояния пациента для данного класса заболеваний, данной нозологии, хорошо известны. Они не раз отмечались в работах ведущих отечественных ученых [3, 4]. Разработка динамической математической модели частной нозологии трудна и довольно трудоемка. Мы попробуем пойти по другому пути. Отечественная медицина в известной степени уже автоматизирована. Современные медицинские информационные системы (МИС) ведут электронные медицинские карты и содержат в различной степени формализованную информацию о миллионах различных клинических случаев. Потенциально мы имеем огромную фактографическую базу данных, отражающую современные медицинские знания, огромное количество прецедентов (реализаций) клинических процессов. Главная идея нашего подхода – это для данного ЛДП на данном шаге процесса найти подходящие прецеденты из множества известных нам клинических процессов. Релевантно отобразить

ные прецеденты могут обогатить врача новыми знаниями, могут помочь врачу принимать клинические решения, формировать лечебно-диагностические действия на данном шаге процесса – в нашей терминологии управлять процессом. Теперь необходимо достичь следующих двух целей: 1) построить математическую модель процесса; 2) предложить метод отбора прецедентов.

2. Математическая модель лечебно-диагностического процесса

Мы будем рассматривать ЛДП как дискретный процесс. Для обозначения шага процесса введем в рассмотрение индексную переменную, значение которой будем указывать верхним индексом, (x^i, u^i) – это ситуация на i -ом шаге процесса.

Отметим важную особенность ЛДП. Управление в текущей ситуации будет определяться не только текущим состоянием, но и управлением на более ранних шагах процесса. Это вытекает из характера самого лечебного процесса. Врач, принимая решение о необходимых в данной ситуации лечебно-диагностических мероприятиях, безусловно, принимает во внимание историю процесса, учитывает уже оказанные на пациента воздействия. Мы имеем дело с управляемым процессом с памятью. В [1, 2] выдвинут важный для построения модели постулат. Было предложено учесть эффект памяти с помощью интегрирования управления. 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. Формально представим управление в следующем виде $u^i = \{(c_{j1}^i, u_{j1}^i), (c_{j2}^i, u_{j2}^i), \dots, (c_{jm}^i, u_{jm}^i)\}$, где c_{jk}^i – это интегральная характеристика элемента управления u_{jk} на i -м шаге процесса. Итак, с нашей точки зрения, управление – это конечная последовательность элементов управления с их интегральными характеристиками. На интегральных характеристиках естественно вводятся операции вычитания (дифференцирования) и сложения (интегрирования).

$$\Delta u^{i-1} = u^i - u^{i-1} = \{(c_{j1}^i - c_{j1}^{i-1}, u_{j1}^i), (c_{j2}^i - c_{j2}^{i-1}, u_{j2}^i), \dots, (c_{jm}^i - c_{jm}^{i-1}, u_{jm}^i)\}.$$

Содержательная интерпретация предложенной модели управления очевидна: u^i – это перечень с интегральными характеристиками всех лечебно-диагностических мероприятий оказанных пациенту к i -му шагу процесса, а Δu^{i-1} – это перечень с интегральными характеристиками лечебно-диагностических мероприятий оказанных пациенту к $i-1$ шагу процесса. Собственно, Δu^{i-1} – это и есть управление на $i-1$ шаге процесса, сопутствующее или способствующее переходу пациента из состояния x^{i-1} в состояние x^i . В [2] показано, как можно элементарно определить операции сложения и вычитания интегральных характеристик управления в самых типичных случаях, когда c_{jk} являются действительными числами, или конечными числовыми интервалами.

Как известно – время лечит. Время, безусловно, является важнейшим лечебным фактором с интегральным характером воздействия. Поэтому время также включается в модель, именно как управление, в своей интегральной характеристике, хотя мы и не можем им явно управлять. Это как бы управление заданное нам самой природой. Под интегральной характеристикой времени в лечебном процессе можно понимать временную длительность, например: число дней прошедших с начала госпитализации, число дней прошедших с момента заболевания, и т.п. Учитывая важность этого фактора, время можно явно выделить в модели, формально описывая ситуацию как (x^i, u^i, t^i) , где t^i – это время, прошедшее от начала процесса к i -му шагу процесса, а $\Delta t^i = t^{i+1} - t^i$ – это длительность i -го шага процесса. Для дальнейшего описания модели нам потребуется привести обширную цитату из работы [2].

«Перейдем от рассмотрения отдельного процесса к рассмотрению ансамбля процессов. Сразу же отметим, что в рамках каждого ансамбля процессы связаны между собой одной нозологией, одним заболеванием. Именно эта связь и позволяет надеяться на то, что, как состояния, так и управления в отдельных процессах будут схожи между собой. Решение диагностической задачи, дифференциальная диагностика в работе не рассматриваются... У нас есть две возможности упростить модельное представление ЛДП в пространстве состояний. Первая возможность связана с тем, что мы рассматриваем процесс в контексте определенной нозологии. Это означает, что диагноз уже поставлен и, фактически, диагностическая задача считается решенной а priori. Возможно, что часть характеристик состояния, которые наблюдались в ходе ЛДП в связи с решением диагностической задачи, можно будет опустить (свернуть) без потери качества управления. Вопрос к врачу о возможности такой свертки звучит так: Когда диагноз уже поставлен, за какими характеристиками состояния пациента можно перестать следить, и при этом не опасаться потерять в качестве ЛДП? Если ответ для какой-то части характеристик состояния будет положительным, то эти характеристики можно будет в модели просто опустить. Вторая возможность свертки связана с введением отношений эквивалентности на характеристиках состояния и, соответственно, с переходом к рассмотрению классов эквивалентных состояний.

Начнем с условного примера. Пусть мы имеем две реализации ЛДП с диагнозом «ОРЗ». Характеристики текущего состояния первого пациента: возраст 39 лет, температура 38,5С. Характеристики текущего состояния второго пациента: возраст 45 лет, температура 38,7С. Пусть, с точки зрения врача, назначающего лечение, различие в значениях двух приведенных характеристик несущественное, и в обоих случаях может быть назначено одно и то же лечение. Объединение этих двух состояний в одно состояние упростит модель и позволит нам считать, что оба процесса находятся в одной ситуации. Для объединения этих состояний мы должны сделать формальный вывод, что с точки зрения данного лечебного процесса, возраст 39 лет «равен», а точнее, в некотором смысле эквивалентен возрасту 45 лет, а температура 38,5°С эквивалентна температуре 38,7°С. И тогда характеристики объединенного состояния можно записать в виде: возраст – «больше или равен 18 годам», температура – [37°С, 39°С], или возраст – «взрослый», температура – «повышенная».

Переходя к более формальным математическим утверждениям, мы говорим о том, что на множестве значений каждой характеристики в рамках данной нозологии можно ввести отношение эквивалентности, которое разобьет множество значений характеристики на классы эквивалентности. Указанные выше в кавычках значения и есть названия соответствующих классов эквивалентности.

Пусть имеются расширенные, включающие управление, состояния ЛДП, которые характеризуются конечными упорядоченными наборами характеристик (признаков). Для упрощения изложения, без потери общности, мы сейчас откажемся от выделения в

расширенном состоянии (x^i, u^i, t^i) управления и времени, и запишем формально расширенное состояние в виде $x = (x_1, x_2, \dots, x_n)$. Информационная природа характеристик (признаков) x_i и их семантика для нас сейчас не важна. Все, что мы потребуем от них – это определенность множества значений каждого признака $x_i \in X_i$ и задание на множестве значений каждого признака отношения эквивалентности \approx . Два состояния процесса A и B будем называть эквивалентными по характеристике (признаку) $j \in \{1, 2, \dots, n\}$, если у наборов признаков соответствующих состояниям $x^A = (x_1^A, x_2^A, \dots, x_n^A)$ и $x^B = (x_1^B, x_2^B, \dots, x_n^B)$ эквивалентны их j -е компоненты: $x_j^A \approx x_j^B$. Два состояния A и B будем называть эквивалентными, если A и B эквивалентны по всем своим признакам: $x_j^A \approx x_j^B, j \in \{1, 2, \dots, n\}$. При этом мы потребуем, чтобы отношения эквивалентности на интегральных характеристиках управления были введены таким образом, чтобы затем на классах эквивалентности состояний были определены и сохраняли семантику ранее введенные операции вычитания (дифференцирования) и сложения (интегрирования) для управления. Удовлетворить это требование будет несложно, так как интегральные характеристики управления зачастую сводятся просто к показателям объема лечебно-диагностических мероприятий. Интегральная числовая характеристика управления для состояния может превратиться в интервальную характеристику для класса, если разбить область определения числовой характеристики на множество непересекающихся и покрывающих область определения характеристики интервалов, и определить отношение эквивалентности через принадлежность к одному интервалу из данного разбиения.

Очевидно, что введенное отношение эквивалентности состояний, позволяет разбивать состояния, для рассматриваемого ансамбля ЛДП, на классы эквивалентности. Заменяя состояния ансамбля ЛДП классами эквивалентности, мы упрощаем модель для данного ансамбля и получаем в модели новое качество. Теперь вместо множества отдельных, не связанных или слабо связанных, в общем случае, реализаций ЛДП, мы можем получить сеть связанных, разделяющих одни и те же эквивалентные состояния, реализаций! Собственно здесь и должна проявиться та самая, ожидаемая из общепсихологических соображений [5, 6], общность ЛДП для данной нозологии. Заметим, что теперь мы можем, и должны! (при адекватной кластеризации состояний), столкнуться с тем, что в один класс мы можем попадать из других различных классов состояний, и переходить из данного класса состояний в другие различные классы состояний, даже при условии одного и того же управления. Таким образом, ансамбль реализаций ЛДП после кластеризации состояний начинает моделироваться случайным дискретным Марковским процессом.

Итак, после введения отношений эквивалентности на характеристиках в расширенном пространстве состояний лечебно-диагностический процесс моделируется дискретным Марковским процессом. На основании экспериментальных данных мы можем дать оценку вероятностям перехода процесса между классами эквивалентных состояний. Для каждого перехода мы можем в модели указать, какие управления вызвали этот переход, указать относительную частоту (оценку вероятности) выбора управления для данного перехода. Для каждого состояния можно также указать все известные модели прецеденты управлений в этом состоянии с оценкой вероятности выбора каждого управления. Марковская модель сразу же открывает перед нами возможности расчета различных статистик ЛДП: вероятности различных исходов процесса, средняя длительность процесса, распределение интегральных характеристик лечебно-диагностических мероприятий (управления). Врач получает возможность, пользуясь данной моделью ЛДП, знать все прецеденты управления для данной ситуации, знать частоту применения в прошлом различных управлений в данной ситуации, прогнозирую-

вать дальнейшее поведение процесса для каждого прецедента управления. В рамках модели можно сформировать желаемую, с медицинской точки зрения, траекторию процесса, для каждого состояния определить желаемые переходы в следующие состояния, определить в каждом состоянии прецеденты управлений, для которых желаемые переходы будут иметь наибольшую вероятность. Модель становится для врача путеводителем по прецедентам лечения данной нозологии, становится своеобразным клиническим руководством по лечению данной нозологии, основанному на реальных фактах, на реальных реализациях ЛДП».

То, что в ходе построения математической модели лечебно-диагностического процесса мы пришли к стохастической модели – весьма отрадный факт. «При детерминированном подходе модельный ЛДП ведет себя полностью детерминировано и определяется начальным состоянием и «управлением» (лечебно-диагностическими мероприятиями), что противоречит общему убеждению о присущей лечебным процессам вариативности и выраженной вероятностной природе. Вряд ли найдутся желающие оспаривать тот факт, что человеческий организм, как систему, следует относить к сложным системам и, согласно классификации систем Р.Эшби, сложные системы и протекающие в них процессы должны иметь стохастическую природу. Итак, одно из принятых нами концептуальных требований к построению модели ЛДП состояло в разработке именно стохастической модели, адекватной природе моделируемых процессов» [1].

Итак, подведем итоги. Мы рассматриваем ансамбль лечебно-диагностических процессов объединенных одной нозологией. Процессы рассматриваются как дискретные. В описании процессов выделяются характеристики управления. На характеристиках управления определяются операции сложения (интегрирования) и вычитания (дифференцирования). Физическое время (индекс шага процесса в модели) рассматривается как управление. На характеристиках вводятся отношения эквивалентности, порождающие классы эквивалентности, как для характеристик, так и для состояний. По ансамблю процессов относительно выделившихся классов состояний строится Марковская модель для данного ансамбля лечебно-диагностических процессов.

3. Меры близости для состояний и процессов

Переходя от рассмотрения отдельных состояний к классам эквивалентных состояний, мы уже фактически решали задачу о близости состояний. Все «достаточно близкие» состояния оказываются в одном классе эквивалентности. Тем не менее, остается актуальным вопрос о близости классов (состояний – представителей классов) друг к другу. Решение этого вопроса имеет важное практическое значение. При появлении, в ходе самообучения модели или в ходе контроля реализации процесса, нового класса эквивалентности, мы имеем дело с новым прецедентом клинической ситуации, ситуации, с которой раньше в модели не сталкивались. Модель не может предложить для новой ситуации прецеденты управления, не может подсказать врачу, что делали ранее в данной ситуации. Очевидным выходом из данной ситуации было бы нахождение в модели состояний близких к данному, близких состояний для которых известны прецеденты управления. Покажем, как на основе введенных в модель отношений эквивалентности на характеристиках расширенного состояния, можно построить метрику в пространстве состояний, и тем самым определить расстояние и количественно оценить близость между состояниями.

Мы продолжим использовать введенные выше обозначения для расширенного состояния $x = (x_1, x_2, \dots, x_n)$. Считаем также, что определены множества значений каждого

признака $x_i \in X_i$ и на множестве значений каждого признака задано отношение эквивалентности \approx .

Интуитивно ясно, что чем большее число компонент двух состояний эквивалентны, тем более схожи между собой эти состояния, тем они ближе друг к другу. Для формализации понятия схожести состояний введем в рассмотрение меры близости состояний и расстояния между состояниями.

Обозначим через N^n множество всех подмножеств конечного множества натуральных чисел $\{1, 2, \dots, n\}$. На N^n введем меру μ – как отображение N^n на множество действительных неотрицательных чисел. Потребуем, чтобы мера μ обладала следующими свойствами:

$$\begin{aligned} \mu(S) &\geq 0, \forall S \in N^n, \\ \text{если } S^1 \supseteq S^2, S^1 \in N^n, S^2 \in N^n, \text{ то } \mu(S^1) &\geq \mu(S^2), \\ \mu(\emptyset) &= 0, \mu(N^n) = n. \end{aligned}$$

Мы потребовали обычных для меры свойств, включая требование нормировки меры. На основании введенной меры определим расстояние между состояниями A и B следующим образом

$$\rho(A, B) = 1 - \frac{1}{n} \mu(S^{AB}),$$

где $S^{AB} \in N^n$ – это множество индексов компонент состояний $x^A = (x_1^A, x_2^A, \dots, x_n^A)$ и $x^B = (x_1^B, x_2^B, \dots, x_n^B)$, по которым A и B эквивалентны:

$$x_j^A \approx x_j^B \quad \forall j \in S^{AB} \text{ и } x_k^A \not\approx x_k^B \quad \forall k \notin S^{AB}, k \in \{1, 2, \dots, n\}.$$

Множитель перед мерой введен для нормировки расстояния, областью определения которого является интервал $[0, 1]$. Расстояние между состояниями полностью эквивалентными по всем своим признакам равно 0. Расстояние между состояниями, не имеющими ни одного эквивалентного признака, равно 1. Удобно также трактовать это расстояние в процентном отношении: состояния подобны на 100% - 0%, при изменении расстояния от 0 до 1. Перейдем к конкретизации возможных мер и расстояний.

Возьмем в качестве меры $\mu(S)$ мощность множества S . Содержательно это означает, что делается предположение, что все признаки «эквивалентны» по своей информативности. Врач, собирающийся использовать эту меру для определения близких состояний, должен это предположение хорошо понимать. В этом случае мера становится аддитивной по признакам, и легко доказать [2], что расстояние становится метрикой. Введем в рассмотрение функцию $g(S^{AB})$ отображающую множество S^{AB} в числовой вектор $\vec{g} = (g_1, g_2, \dots, g_n)$ с компонентами, принимающими значения 0 и 1. Компонента вектора \vec{g} равна 1, если индекс этой компоненты входит в множество S^{AB} , и нулю в противном случае. Тогда можно переписать выражение для расстояния в следующем виде

$$\rho(A, B) = 1 - \frac{1}{n} \sum_{i=1}^n g_i^{AB}.$$

Легко заметить, что можно также определить метрику между состояниями как взвешенную сумму компонентов вектора $\vec{g} = (g_1, g_2, \dots, g_n)$.

$$\rho(A, B) = 1 - \frac{1}{m} \sum_{i=1}^n g_i^{AB} \times c_i,$$

где $\vec{c} = (c_1, c_2, \dots, c_n)$ – вектор весов с положительными компонентами, а $m = \sum_{i=1}^n c_i$ – нор-

мировочный множитель. Более общее определение расстояния, позволяет учитывать относительную информативность признаков. Содержательно определить веса признаков может врач-эксперт, у него появляется возможность выделить более значимые признаки и повысить их вклад в метрику.

Часто некоторые группы признаков демонстрируют синергетический эффект взаимодействия, выражающийся в известном философском изречении: «целое – больше суммы своих частей». Это означает, что мера близости в общем случае должна быть неаддитивной относительно признаков. Поэтому, анализируя уже содержательную природу и информативность как отдельных признаков, так и их групп в целом, мы можем пытаться явно учесть этот системный эффект и явно определять меру μ на множестве N^n . Однако вопрос о том, будет ли для произвольной меры введенное расстояние метрикой, остается для нас открытым. Да и само содержательное определение такой меры может дать только врач-эксперт.

Практическое значение от введения расстояния между состояниями мы уже отметили. Врач, оказавшись в новой для себя ситуации, формально может определить самые близкие ранее наблюдавшиеся известные модели состояния и рассмотреть течение процесса в этих состояниях. Отметим, что кластеризация состояний и предложенная количественная мера близости состояний взаимно дополняют друг друга. При «слабых» отношениях эквивалентности и соответственно «сильной» кластеризации состояний необходимость в нахождении близких состояний будет меньше, чем в условиях, когда определены «сильные» отношения эквивалентности и, соответственно, число классов велико и велика вероятность наблюдения нового ранее неизвестного класса. В предельном случае, определив отношение эквивалентности для каждой характеристики как равенство, мы приходим к множеству «непредставительных», т.е. представленных в пределе всего одним состоянием, классов. В этом предельном случае применение мер близости состояний даст врачу инструмент поиска общности и генерализации опытных данных.

4. Численное моделирование ЛДП по реальным клиническим данным

Силами сотрудников и аспирантов 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 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. Авторы работы благодарят врача-терапевта к.м.н. Фролову Маргариту Александровну за ее участие в работе по нормализации.

Были рассмотрены три ансамбля процессов, см. таблицу 1.

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	48909/48513	<3%	871/99

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 случае 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 2 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. Так же 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, т.е. 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 незначи-

тельным при малом числе процессов в моделируемом ансамбле, в этом случае индивидуальные черты процесса будут отделять его от других непохожих на него. К этому случаю можно отнести нозологию J13, всего 166 процессов, менее 1% сжатия. По мере увеличения мощности моделируемого ансамбля процессы 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 шириной нозологии.

Таблица 2. Упорядоченный по убыванию частоты словарь характеристик расширенного состояния пациента для нозологии «Пневмония, вызванная *Streptococcus pneumoniae*».

№	Рейтинг характеристики (число процессов, в которых эта характеристика фиксировалась)	Название характеристики
1	2938	Возраст
2	2938	Пол
3	2688	ЭКГ
4	2667	Фибриноген
5	2623	Креатинин в сыворотке крови
6	2612	Индивидуальные занятия ЛГ при заболевании легких
7	2588	Глюкоза крови
8	2540	Мочи анализ общий
9	2504	АСТ активность
10	2496	АЛТ активность
11	2350	Крови анализ развернутый

5. Заключение

В работе рассмотрена математическая модель лечебно-диагностического процесса, базирующаяся на прецедентном подходе. Из БД МИС извлекается накопленное в ней и формализованное знание в форме прецедентов ЛДП по данной нозологии. ЛДП рассматривается как дискретный управляемый Марковский процесс с памятью. Состояния, они же прецеденты, являются результатом статистической обработки и генерализации знаний, заключенных в ансамбле реализаций ЛДП для данной нозологии. Эффект памяти для управляемого процесса предлагается учесть в виде интегральных характеристик управления, включаемых в расширенное состояние процесса. Кластеризация в пространстве состояний производится на основе введения в модель отношений эквивалентности на характеристиках состояния. Включение в модель отношений эквивалентности позволяет построить метрику в пространстве состояний и оценивать близость состояний ЛДП друг к другу. Сама метрика параметризуется мерой, задаваемой относительно характеристик состояния. ЛДП рассматривается в контексте уже решенной диагностической задачи, что значительно упрощает моделирование процесса. Приведены первые результаты численного моделирования ЛДП по реальным клиническим данным, подтверждающие перспективность предложенной математической модели. Результаты работы могут найти широкое практическое применение в МИС пятого поколения, представляют интерес для разработчиков медицинских информационных систем и для медиков, решающих задачи управления качеством и совершенствования ЛДП.

Работа выполнена при поддержке Российского фонда фундаментальных исследований (13-07-12012).

Список литературы

1. 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.
2. Малых В.Л., Гулиев Я.И. Управляемый стохастический прецедентный процесс с памятью как математическая модель лечебно-диагностического процесса // Информационные технологии и вычислительные системы. 2014. № 1.
3. Гельфанд И.М., Розенфельд Б.И., Шифрин М.А. Очерки о совместной работе математиков и врачей. М.: Едиториал УРСС, 2005.
4. Назаренко Г.И., Осипов Г.С. Основы теории медицинских технологических процессов. Т. 1. М.: Физматлит, 2005.
5. Малых В.Л., Гулиев Я.И. Прецеденты в медицинских информационных системах // Программные продукты и системы. 2009. № 2 (86). С.19-27.
6. Белов А.К. Философия естественной природы. (Метафизика для физиков и математиков). Часть первая. Совершенная материальная действительность. М.: Издательство «Спутник+», 2011.