

Решение задачи запроса дополнительной информации на основе декларативной темпоральной базы знаний для дифференциальной диагностики

Клещев А.С., Шалфеева Е.А.

Институт автоматики и процессов управления, ул. Радио, 5, г. Владивосток, Россия.

kleshev@iacp.dvo.ru, shalf@iacp.dvo.ru

Аннотация. В предметной области «медицина» построено большое множество экспертных систем как для решения отдельных задач, так и для нескольких. Как правило, ЭС поддерживают задачу диагностики, предлагая специалисту множество полезных гипотез, такое расширенное множество гипотез далее должно быть уменьшено. По результатам наблюдений ситуации, для которых существует более одной гипотезы о классе, требуется предложить дополнительное наблюдение ситуации, которое позволит сократить множество гипотез. Однако методам решения важной подзадачи запроса дополнительной информации для сокращения множества гипотез в литературе уделено недостаточно внимания. Известные методы решения запроса дополнительной информации для сокращения множества гипотез о диагнозе не учитывают то обстоятельство, что болезнь является процессом, протекающим во времени, и проявления болезни также могут изменяться с течением времени. Предлагаемый в работе метод алгоритм ранжирования дифференциаторов для решения такой задачи позволяет снять указанные ограничения.

Ключевые слова: гипотезы о диагнозе, результаты наблюдения, запрос дополнительной информации, база знаний, дифференциатор

1 Введение

Медицина - типичный пример предметной области с множеством сложных и ответственных задач принятия разных типов решений. После решения о первичном сборе информации о пациенте врач принимает решения о его диагнозе, его лечении и прогнозе изменения состояния пациента в результате лечебных мероприятий. На основе этого прогноза планируется дальнейшее наблюдение за ним. Если результаты такого наблюдения не соответствуют прогнозу, принимаются решения о коррекции диагноза и/или плана лечения, что ведет к корректировке прогноза и т.д.

В предметной области «медицина» построено большое множество экспертных систем (как исследовательских прототипов) для поддержки деятельности врачей. Есть экспертные системы (ЭС) как для решения отдельных задач, так и для нескольких. Некоторые специалисты ожидают от ЭС поддержки в процессе выдвижения множества полезных гипотез, некоторые - в процессе выдвижения одной наиболее вероятной гипотезы, некоторым необходима конструктивная критика выдвинутой гипотезы. Как правило, ЭС поддерживают задачу диагностики, предлагая специалисту множество полезных гипотез, такое расширенное множество гипотез далее должно быть уменьшено. Однако методам решения важной подзадачи запроса дополнительной информации для сокращения множества гипотез в литературе уделено недостаточно внимания.

2 Теоретические проблемы

2.1 Постановка задачи

При решении задачи диагностики врач вынужден одновременно решать и другую задачу – сбора необходимой информации о больном, а именно, он должен решать какие исследования должен выполнить он сам и другие специалисты, чтобы на основе собранной информации поставить правильный диагноз больному. Экспертные системы, решающие задачу медицинской диагностики, можно разделить на два класса: которые полностью перекалывают на врача решение задачи сбора необходимой информации и решают лишь задачу медицинской диагностики на основе введенных врачом данных о больном [1,2,3,4,5]; которые помимо решения задачи медицинской диагностики дают врачу рекомендации относительно того, какая информация о больном необходима для этого [6,7,8,9,10,11,12,13]. В свою очередь, ЭС второго класса можно разделить на два подкласса: ЭС, в которых задача запроса информации о больном не отделяется от задачи медицинской диагностики, а решается в ходе выполнения общего для обеих задач алгоритма [10,11]; ЭС, в которых эти две задачи разделены [6-14]. Информация, получаемая врачом при обследовании больного, используется им как для выдвижения такого множества гипотез о диагнозе, в котором должна содержаться правильная гипотеза, так и для сокращения этого множества до одной гипотезы. Среди ЭС, в которых задача запроса информации отделяется от задачи диагностики, можно выделить два пересекающихся подкласса: ЭС, в которых запрашивается информация для выдвижения гипотез о диагнозе [7]; ЭС, в которых запрашивается информация для сокращения множества гипотез [6-13].

Среди ЭС, в которых запрашивается информация для сокращения множества гипотез, можно выделить:

- 1) ЭС, в которых предполагается, что диагноз пациента известен, но требуется дополнительная информация для диагностики стадии развития заболевания [7];
- 2) ЭС, в которых каждая гипотеза о диагнозе содержит несколько заболеваний (либо основное заболевание и его осложнения, либо, когда имеет место сочетанная патология) [7,8];
- 3) ЭС, в которых каждая гипотеза о диагнозе содержит только одно заболевание [6,10,11,13].

Далее рассматривается лишь последний случай. Формальная постановка задачи запроса дополнительной информации [15] в этом случае состоит в следующем.

Дано:

База Знаний (KB), удовлетворяющая условию разделимости классов, результаты наблюдения ситуации и характеристики системы\пациента (R), такие что для них множество гипотез (HR_{KB}) имеет мощность не меньше двух.

Найти: такой запрос дополнительной информации (Q) о ситуации\системе для результатов R, согласованных с онтологией, что для R, объединенного с ответом (AQ) на этот запрос, новое множество гипотез имеет меньшую мощность, чем множество гипотез HR_{KB} .

Таким образом, суть задачи запроса дополнительной информации для сокращения множества гипотез о диагнозе состоит в следующем: по результатам наблюдений ситуации, для которых существует более одной гипотезы о классе, требуется предложить дополнительное наблюдение ситуации, которое позволит сократить множество гипотез.

2.2 Известные методы решения

В литературе рассматриваются следующие методы запроса дополнительной информации для сокращения множества гипотез о диагнозе:

- 1) в первую очередь запрашиваются значения таких признаков, что какое бы значение ни было получено в результате наблюдения любого такого признака, множество гипотез о диагнозе всегда будет сокращено [10,11,13];
- 2) затем запрашиваются значения таких признаков, что при получении некоторых значений такого признака множество гипотез о диагнозе может быть сокращено; при ранжировании таких

признаков должны учитываться априорные вероятности гипотез, а также вероятности появления различных значений этих признаков для рассматриваемых заболеваний;

3) могут запрашиваться значения нескольких признаков одновременно, если в результате их наблюдения могут быть получены такие комбинации их значений, при которых множество гипотез о диагнозе может быть сокращено («синергетические эффекты», когда два признака должны присутствовать [16], или синдромы).

Такие методы применимы при однократном обследовании больного и не учитывают промежутка времени, прошедшего с момента начала заболевания до момента обследования.

Описанные конкретные подходы к выбору методов решения задачи запроса дополнительной информации для распознавания и Методы решения задачи опираются на *дополнительную информацию* в БЗ, и имеются «вариации» содержания этой *дополнительной информации* в БЗ..

Вариация 1а:

В Знания (БЗ) вводится *классификация* заболеваний, исходя из их симптоматики; у классов заболеваний есть (общие) признаки. Примеры: иерархическая классификация типов заболеваний, когда все болезни устроены в иерархию (дерево заболеваний), организованную вокруг систем органов. Каждый нетерминал в этой иерархии связан с проявлениями, которые характерны для всех ее подчиненных [9], классы состояний верхнего уровня: уровня общности, для которого определяется наличие или отсутствие общих патофизиологических процессов по отношению к частным патофизиологическим процессам, специфичным для нижнего уровня [7]. Это позволит сначала дифференцировать классы, и затем, когда гипотез-классов заболеваний останется мало, свести задачу к задаче запроса дополнительной информации.

Вариация 1б:

В Знания (БЗ) вводится иерархия стадий рассуждения при анализе заболеваний.

Случай 1. корнем «дерева» является главное проявление (примеры: «слои» предопределенного дерева («Differential-Diagnosis set») «подражают» гипотетико-дедуктивному диагностическому процессу рассуждения клинициста, корнем дерева является главное проявление, которое вызвало диагностический процесс [16], «под» ним определены подгруппы диагнозов с разными вариантами главного проявления, «под» каждым из которых определены меньшие (по числу) подгруппы диагнозов с добавлением дифференцирующего и т.д., пока дифференцирующий признак не обеспечит свидетельство для одного (или нескольких) конкретных заболеваний [16]).

Случай 2. корнем дерева является обобщенная причина (пример: на верхнем уровне дерева - инфекции - «significant infection which should be treated» (с рекомендуемыми курсами лекарственной терапии), подцели - культуры; микроорганизмы, многие из которых распадаются на более мелкие подцели, а листьями дерева являются лабораторные факты [10,11].

Вариация 2:

в Знания (БЗ) вводятся связи между заболеванием и синдромом или несколькими признаками [16]. При этом от признаков иногда может требоваться, чтобы два признака проявлялись в определенном порядке, или чтобы событие предшествовало внешнему признаку, в том числе с нужным интервалом (пример: время принятия внутрь подозрительной еды равняется 3 ± 3 ч до времени начала диареи) [16]. Имеется и такая разновидность: вводятся связи между заболеванием и его достаточным симптомокомплексом (из так называемого «клинического конверта»). При этом симптомокомплексов может быть несколько. Такая дополнительная информация в Знаниях чаще нужна для выдвижения наиболее вероятной гипотезы, реже – для множества полезных гипотез.

При выборе порядка запроса значений признаков должны также учитываться стоимость проведения исследований и степень их травматичности для больного. Технология разработки стандартов (клинических протоколов) предусматривает рассмотрение уровней доказательности и убедительности признаков [17].

Методы, приведенные выше, применимы при однократном обследовании больного и не учитывают то обстоятельство, что болезнь является процессом, протекающим во времени, и проявления болезни также могут изменяться с течением времени. Чтобы обойти эту проблему, в литературе иногда в качестве значений признаков рассматриваются продолжительность присутствия тех или иных значений этих признаков, что решает проблему лишь частично. Кроме того, эти методы

не учитывают влияние событий (лечебных мероприятий и т.п.), связанных с больным, на значения и их изменение во времени наблюдаемых у больного признаков. Наконец, как правило, эти методы не учитывают влияния индивидуальных характеристик больного (пола, возраста и т.п.) на проявления заболеваний.

2.3 Постановка задачи запроса дополнительной информации для предметной области, в которой процессы связаны между собой посредством причинно-следственных отношений

В ряде предметных областей решаются задачи относительно систем, в которых во времени протекают процессы, связанные между собой посредством причинно-следственных отношений. Процессы, протекающие в такой системе, условно можно разделить на внешние (наблюдаемые) и внутренние (которые не могут наблюдаться непосредственно и о которых можно судить лишь по их связям с наблюдаемыми процессами). Наблюдаемые процессы обычно называют признаками, признаки имеют значения, которые получаются в результате наблюдения этих признаков и изменяются с течением времени (в онтологии содержатся термины, денотаты которых являются функциональными соответствиями, зависящими от времени). Помимо признаков наблюдаемыми могут быть постоянные во времени характеристики системы и происходящие в различные моменты времени события, внешние по отношению к системе, но воздействующие на протекающие в ней процессы (как внешние, так и внутренние) [15]. Онтология любой предметной области состоит из *терминологии* и множества предложений (аксиом), представляющих те свойства терминов онтологии, которые определяются соглашениями. Терминология предметной области включает группы терминов, интерпретациями которых являются значения, функциональные соответствия и отношения. Функции и отношения, являющиеся интерпретациями функциональных и предикатных символов, могут зависеть от времени, координат пространства и других характеристик [15].

Причинно-следственные отношения между множеством причин и множеством следствий включают: отношения между внутренними и внешними процессами, отношения между событиями и внешними процессами, отношения между событиями и внутренними процессами и т.д. [15]. Когда требуется распознать внутренний процесс, не присущий «системе» (отклонение от нормы или, например, заболевание, изначально не присущее здоровому организму), то говорят о задачах диагностики.

Обозначим внутренние процессы как F_{in} (множество функциональных символов, зависящих от времени, значения которых - состояния процесса); упорядоченное множество моментов времени - $\{t_0, \dots, t_k\}$, внешние процессы - F_{ex} (наблюдаемые в разные моменты времени значения признаков - $R_{ex}(t_0, \dots, t_k)$); характеристики системы - R_0 ; события - F_{ev} (наблюдаемые в некоторые моменты времени - $R_{ev}(t_j)$). Предикаты, аргументами которых помимо моментов времени являются причины, воздействующие факторы и следствия, обозначим P_c .

При диагностике учитывается промежуток времени, прошедший с момента начала развития внутреннего процесса (не присущего «системе»), влияние событий на значения признаков и их изменение во времени, влияние индивидуальных характеристик «системы» на проявления нарушений. Обозначим внутренние процессы, не присущие системе, как $\Delta \subseteq F_i$.

В темпоральную базу знаний KB входят предложения о свойствах ситуаций каждого класса Δ_i из Δ . В общем случае среди этих предложений есть:

варианты развития, связывающие значения некоторого события с диагнозом и зависящие, возможно, от значений воздействующих факторов;

варианты развития, связывающие диагноз с изменениями значений признака и зависящие от периода развития внутреннего процесса и от значений воздействующих факторов;

варианты развития проявления, измененного воздействием события, связывающие диагноз и событие с некоторым значением признака и зависящие, возможно, от значений воздействующих факторов;

варианты развития реакции на воздействие события, связывающие значение некоторого события с некоторым значением признака и зависящие, возможно, от значений воздействующих факторов [18].

Замечание. Те варианты развития клинического проявления или варианты развития реакции на воздействие события, в которых для исследования признаков существуют несколько методов, позволяющих получить одно и то же свидетельство (значение признака), в БЗ должны быть сгруппированы по признакам (а не по методам получения значений).

Для соответствия вышеупомянутой современной технологии разработки клинических протоколов в КВ в предложениях, связывающих заболевание со значениями признака, для признаков могут быть указаны уровни доказательности и убедительности.

Пример шкалы значений (оценки) убедительности доказательств: А (доказательства убедительны), В (относительная убедительность доказательств), С (остаточных доказательств нет), D («отрицательные» свидетельства \ доказательства).

При получении при диагностике n гипотез (о диагнозе) требуется принятие решения о запросе дополнительной информации (о дополнительном обследовании) для сокращения множества гипотез о диагнозе. В этих условиях задача запроса дополнительной информации, учитывающая промежуток времени, прошедший с момента начала развития диагностируемого внутреннего процесса до момента принятия решения (решения о запросе дополнительной информации), такова.

Дано:

множество гипотез $H_{R,KB} = \{\Delta_1, \dots, \Delta_m\}$, имеющее мощность не меньше двух;

результаты наблюдения пациента (ситуации) в разные моменты времени и постоянные характеристики $R = R_0 \cup R_{ex}(t_0, \dots, t_k) \cup R_{ev}(t_j)$, такие что для них выдвинуто множество гипотез ($H_{R,KB}$);

темпоральная БЗ КВ, удовлетворяющая условию разделимости классов (диагнозов $\Delta_1, \dots, \Delta_m$).

Найти: такой запрос дополнительной информации Q о пациенте\ситуации для результатов R , согласованных с онтологией медицинской диагностики, что для R , объединенного с ответом AQ на этот запрос, новое множество гипотез имеет меньшую мощность, чем множество гипотез $H_{R,KB}$ (т.е. хотя бы одно предложение из БЗ, истинное для причинно-следственной модели пациента\ситуации $P_c(R, \Delta_{ij})$, становится ложным для причинно-следственной модели $P_c(R \cup AQ, \Delta_{ij})$).

3 Метод ранжирования дифференциаторов

Предлагаемый метод решения задачи запроса дополнительной информации для сокращения множества гипотез о диагнозе позволяет снять указанные выше ограничения.

Для удобства предлагается среди запрашиваемых признаков различать дифференциаторы множества рассматриваемых гипотез и антидифференциаторы (рассматриваемого множества). Дифференциатор – тот признак, каждое значение которого сокращает множество гипотез; антидифференциатор – такой признак, ни одно из значений которого НЕ сокращает множество гипотез; «частичные» дифференциаторы - остальные признаки, которые могут быть ранжированы (по доле значений, которые сократят множество гипотез).

3.1 Алгоритм ранжирования дифференциаторов по времени, влиянию событий и индивидуальных характеристик системы

Для решения такой задачи запроса дополнительной информации с использованием знаний о внутренних процессах и влиянии на них внешних событий предлагаются следующие этапы алгоритма выявления среди известных измеряемых признаков дифференциаторов и «частичных» дифференциаторов.

Шаг 1) Предобработка БЗ и построение «рабочей» модели признаков:

Из БЗ о заболеваниях КВ выделить фрагмент КВ', содержащий предложения о свойствах ситуаций (пациентов) каждого класса (диагноза) Δ_{ij} из множества гипотез $H_{R,KB} = \{\Delta_1, \dots, \Delta_m\}$;

инвертировать фрагмент КВ', т.е. построить «рабочую» модель признаков КВ'^{inv}: чтобы в модели рассматривать только подмножество признаков $f_{ex} \in F_{ex}$ и событий $f_{ev} \in F_{ev}$, которые *связаны с гипотезами* из $H_{R,KB}$.

Шаг 2) Исключение лишних признаков из «рабочей» модели:

Обозначим $t_{\text{мом}}$ текущий момент - точку на временной оси от начала заболевания до момента принятия решения о запросе дополнительной информации. Назовем «необходимыми условиями» индивидуальные характеристики пациентов ($\in R_O$), или события ($\in R_{ev}$); содержащиеся в предложении для класса - диагноза Δ_{ij}

В «рабочей» модели удалить (или пометить как «неактивные»):

каждый признак $f_{ex}Name_i$, который:

является статическим и не требующим выполнения «необходимых условий» и его значение $f_{ex}Value_i$ уже присутствует в R ;

каждый статический признак $f_{ex}Name_i$, который:

отсутствует в $R_{ex}(t_0, \dots, t_k)$, но

для которого известно, что не выполнены «необходимые условия» (НУ);

те признаки, которые:

имеют единственный вариант динамики значений во все последующие моменты времени, начиная с «текущего момента», и для которого не выполнены НУ;

те варианты динамики значений $f_{ex}Value_i(t_{\text{мом}}, \dots, t_k)$, которые:

продолжаются с «текущего момента» и

для которых из $R_O \cup R_{ev}$ известно, что не выполнены НУ;

а также те признаки $f_{ex}Name_i$, у которых степень убедительности доказательств = «Е» (Веские отрицательные доказательства).

Примечание. При реализации алгоритма на шаге 1 в «рабочей» модели пометить как «неизвестный» (или удалить): тот признак $f_{ex}Name_i$, который был запрошен при предыдущей попытке дифференциации, но ответ на него не поступил.

Остальные считать *возможными дифференциаторами*.

Шаг 3) найти и исключить антидифференциаторы:

Те признаки в KB^{Inv} , у которых:

есть связь с каждой гипотезой: $\Delta_{i1}, \dots, \Delta_{im}$;

при этом:

либо есть связь одного и того же значения (статического) признака с каждой гипотезой: $\Delta_{i1}, \dots, \Delta_{im}$ (при отсутствующих либо одинаковых НУ),

либо есть связь одного и того же множества вариантов значений (одного и того же диапазона значений) с каждой гипотезой: $\Delta_{i1}, \dots, \Delta_{im}$ (при отсутствующих либо одинаковых НУ),

есть связь одного и того же множества вариантов значений (одного и того же диапазона значений) с каждой гипотезой: $\Delta_{i1}, \dots, \Delta_{im}$ (при отсутствующих либо одинаковых НУ),

либо, начиная с «текущего момента» (точки на временной оси от начала заболевания), с каждой гипотезой: $\Delta_{i1}, \dots, \Delta_{im}$ связаны одинаковое значение признака или одинаковый диапазон значений или одинаковые варианты динамики значений во все последующие моменты времени,

либо с каждой гипотезой: $\Delta_{i1}, \dots, \Delta_{im}$ связаны вообще одинаковые варианты динамики значений (при выполненных НУ),

считать антидифференциаторами. Их можно удалить из KB^{Inv} («рабочей» модели признаков заболеваний-гипотез). Примечание. Если все признаки KB^{Inv} оказались антидифференциаторами, важно сигнализировать о неполноте БЗ и завершить алгоритм.

Ко множеству антидифференциаторов могут быть отнесены те признаки, у которых убедительность доказательств везде в «рабочей» модели ниже «С».

Шаг 4) искать дифференциатор:

Примечание. В KB^{Inv} содержатся признаки, которые у пациента еще не имеют значений (отсутствуют в $R_{ex}(t_0, \dots, t_k)$), для которых либо отсутствуют НУ, либо выполнены их НУ, либо пока не известно, что они не выполнены.

В KB^{Inv} найти признак $f_{ex}Name_i$, который для всех гипотез имеет разное значение из допустимого множества значений.

А) искать дифференциатор - статический признак с «качественными» значениями:

Среди признаков $\{f_{ex}Name_i\}$ (в KB^{Inv}), имеющих область возможных значений (ОВЗ) – конечное множество значений (перечисляемый тип), искать $f_{ex}Name_i$, который для всех n гипотез имеет разное значение из допустимого множества значений $\{f_{ex}Value_{ij}\}$ $j=1, n$.

Для признаков $\{f_{ex}Name_i\}$ (в KB^{Inv}), присутствующих в R ($R_{ex}(t_0, \dots, t_k)$), но требующих выполнения НУ в виде информации о произошедших событиях, искать те возможные события $f_{ev}Name_w$ ($\in F_{ev}$), запрос Q о существовании (или значении) которых в момент $t_v \leq t_0$ может иметь ответ (AQ) в t_{mom} . Для событий AQ обычно имеет ОВЗ {да, нет}.

Б) искать дифференциатор - статический признак с диапазонами значений:

Среди статических признаков $\{f_{ex}Name_i\}$ (в KB^{Inv}), имеющих в качестве ОВЗ один или совокупность числовых диапазонов,

найти признак $f_{ex}Name_i$, который для всех гипотез имеет непересекающийся поддиапазон значений (от $minValue_j$ до $maxValue_j$), т.е. $minValue_{j+1} > maxValue_j$, $minValue_j > maxValue_{j-1}$.

В) искать дифференциатор - динамический признак:

Среди признаков $\{f_{ex}Name_i\}$, имеющих в любой момент (или период) времени в качестве ОВЗ множество значений, искать $f_{ex}Name_i$, который

в некоторый момент t_v (или период времени $\langle t_v, t_{v+1} \rangle$), не раньше t_{mom} ,

для всех n гипотез имеет разное значение $f_{ex}Value_i(t_v)$ из допустимого множества значений $\{f_{ex}Value_{ij}\}$.

А среди признаков $\{f_{ex}Name_i\}$, имеющих в любой момент (или период) времени в качестве ОВЗ числовые диапазоны,

искать признак $f_{ex}Name_i$, который

в некоторый момент t_v (или период времени $\langle t_v, t_{v+1} \rangle$), не раньше t_{mom} ,

для каждой из n гипотез имеет такой поддиапазон значений ($minValue_j - maxValue_j$), $j=1, n$, что множество таких n поддиапазонов не пересекаются.

Г) оценить найденные дифференциаторы:

Среди дифференциаторов $\{f_{ex}Name_i\}$ (в некоторый момент t_v времени), выбрать те, про которые известно, что на их значение влияют события $f_{ev}Name_w$ (например, приём внутрь ЛС или других веществ) и известен временной интервал начала такого воздействия.

Если для каждого такого «подверженного влиянию» дифференциатора $f_{ex}Name_i$ ответы (AQ) на запросы Q о таких событиях в моменты $t_v \leq t_{mom}$ отсутствуют в $R_{ex}(t_0, \dots, t_k)$, то $f_{ex}Name_i$ считать дифференциатором,

во-первых, в паре с AQ,

во-вторых, при выполнении условий: либо если AQ = нет; либо если AQ = да, но $f_{ex}Value_i(t_w)$ в $t_w \geq t_{mom}$, даже с учётом влияния события имеет для всех гипотез разное значение (или поддиапазон значений) из ОВЗ.

Не «подверженные влиянию» потенциальные дифференциаторы $f_{ex}Name_i$ также считать дифференциаторами.

Шаг 4 упрощается при отсутствии в рабочей модели KB^{Inv} признаков с динамическими значениями. В рабочей модели KB^{Inv} описание связей признаков с заболеваниями может быть

сгруппировано по Вариантам значений. Найти дифференциатор значит попытаться найти $f_{ex}Name_i$, который для всех n гипотез имеет разное значение из допустимого множества значений $\{f_{ex}Value_{ij}\}$ $j=1, n$.

Примечание. Ко множеству дифференциаторов не могут быть отнесены те признаки, у которых убедительность доказательств везде в «рабочей» модели ниже «В».

Если множество дифференциаторов $\neq \emptyset$, то важно рассмотреть для каждого из них степень убедительности доказательств (признака) и убедиться в том, что для каждого из них оценка убедительности доказательств = «А» (доказательства убедительны) и уровень достоверности доказательств = I.

Если множество дифференциаторов = \emptyset , применить шаг 5.

Шаг 5) ранжировать признаки («частичные» дифференциаторы) в модели признаков заболеваний;

Примечание. После шага 4 в KB^{Inv} содержатся частичные дифференциаторы, которые могут дифференцировать гипотезы, если ответом (AQ) станет «подходящее» \ «дифференцирующее» (способное отвергнуть некоторую гипотезу) значение.

При наличии в рабочей модели KB^{Inv} признаков с динамическими значениями признаки в KB^{Inv} могут быть сгруппированы по их Вариантам динамики;

кроме того, в KB^{Inv} значения признаков могут быть сгруппированы по периодам динамики (которым сопоставлены заболевания, при которых такой вариант возможен);

кроме того, рабочая модель KB^{Inv} может быть организована по значениям признаков в разные подынтервалы периодов динамики).

Используя KB^{Inv} ранжирование признаков $f_{ex}Name_i$ можно провести по принципу: какова доля значений, которые сократят множество гипотез (с какой вероятностью выпадет в ответе AQ значение $f_{ex}Value_{ij}$; способное исключить k из n гипотез).

Шаг 5 упрощается при отсутствии в рабочей модели KB^{Inv} признаков с динамическими значениями. В KB^{Inv} описание связей признаков с заболеваниями может быть сгруппировано по Вариантам значений.

Ранжирование статических признаков с качественными значениями (признаки $\{f_{ex}Name_i\}$, имеющие ОВЗ - множество «качественных» значений) таково:

признаки $\{f_{ex}Name_i\}$ (в KB^{Inv}), имеющие ОВЗ множество «качественных» значений, связанные более, чем с одним заболеванием, упорядочить по величине отношения числа разных значений $f_{ex}Value_{ij}$ признака, связанных только с одним из этих заболеваний, к числу *возможных значений*. Далее упорядочить аналогично признаки, у которых разные значения связаны только с двумя заболеваниями (если число заболеваний в рабочей модели >2) и т.д.

Ранжирование статических признаков $\{f_{ex}Name_i\}$, имеющих ОВЗ - множество числовых интервалов, таково:

все такие признаки $\{f_{ex}Name_i\}$, связанные более, чем с одним заболеванием, упорядочить по величине отношения длины интервала ($f_{ex}maxValue_{ij} - f_{ex}minValue_{ij}$), связанного только с одним из этих заболеваний, к длине ОВЗ ($f_{ex}maxValue - f_{ex}minValue$). алгоритма выявления среди известных измеряемых признаков дифференциаторов и «частичных» дифференциаторов.

4 Заключение

Представленный алгоритм выявления дифференциаторов среди известных признаков (в соответствии с постановкой задачи запроса) применяется многократно в процессе дифференциации. При наличии n гипотез о диагнозе-заболевании обращение к алгоритму даст ответ на вопрос «Какой признак запросить, чтобы ответ на него помог отвергнуть хотя бы одну гипотезу?». Если после работы алгоритма запрос признака привел к получению ответа (значение указанного признака в указанный момент времени), то этим ответом либо в соответствии с рекомендацией алгоритма может быть отвергнута указанная гипотеза либо, поскольку представление о состоянии пациента уточняется, запускается процесс генерации нового множества гипотез, ожидаемо меньшего. В результате

оказывается $n' = n-1$ гипотез (или даже меньше). Если $n' > 1$, то необходимо очередное обращение к этому алгоритму. В идеале цикл обращений завершается, когда $n' = 1$.

Описанный метод запроса дополнительной информации может быть применён не только для сокращения множества гипотез о диагнозе, но и для оптимизации процесса «квалификации» тяжести клинической ситуации, требующей немедленных мер (например, вызову СМП).

5 Благодарности

Работа выполнена при частичной финансовой поддержке грантов РФФИ № 15-07-03193 и 16-07-00340.

Литература

- [1] Жмудяк М.Л., Повалихин А.Н., Стребуков, А.В. и др. Диагностика заболеваний методами теории вероятностей. *Изд-во АлтГТУ, Алт. гос. тех. ун-т им. И.И. Ползунова*. 168 с. Барнаул 2006.
- [2] Черняховская М.Ю. Оценка ЭС медицинской диагностики "Консультант-2" на архивном материале нескольких клиник. *ИАПУ ДВО РАН*. 30 с. Владивосток 1989.
- [3] Москаленко Ф.М. Алгоритм диагностики, основанный на реальной онтологии медицины, для многопроцессорной ЭВМ. *Доклад III Междун. конф. "Расо 2006", Институт проблем управления им. В.А. Трапезникова РАН*, Москва 2006.
- [4] Long, W.: Temporal reasoning for diagnosis in a causal probabilistic knowledge base. *Artif Intell Med* 8(3): 193-215 (1996).
- [5] Wolpaw T, Papp KK, Bordage G. Using SNAPPS to facilitate the expression of clinical reasoning and uncertainties: a randomized comparison group trial. *Acad Med*-84(4):517-524? 2009.
- [6] Ле Н. В., Камаев В. А., Панченко Д. П., Трушкина О. А. Модель представления знаний при создании медицинской экспертной системы дифференциальной диагностики. *Известия Волгоградского государственного технического университета* – 6: 42-50, 2014
- [7] Литвиненко В.И., Соколова Н.А., Бень А.П. Некоторые аспекты категорного представления информации в лечебно-диагностических системах. *Вестник ХГТУ* – 1: 101 -104, 1997.
- [8] Lemaire J.B., Schaefer J.P., Martin L.A., et al. Effectiveness of the Quick Medical Reference as a diagnostic tool. *CMAJ* - 161 (6): 725-728, 1999.
- [9] Miller R.A., Pople H., and Myers J. INTERNIST-1: An Experimental Computer-Based Diagnostic Consultant for General Internal Medicine. *N Engl J Med (New England Journal of Medicine)* – 307: 468-476, 1982.
- [10] Patil R.S., Szolovits P., Schwartz W.B. Modelling Knowledge of the Patient in Aced-base and Electrolyte Disorders. *Artificial Intelegence in Medicine*, Boulder, CO: Westview Press, 191-226, 1982.
- [11] Patil R.S., Ramesh S. Causal Representation of Patient Illness for Electrolyte and Acid-Base Diagnosis. *MIT Lab for Comp. Sci. TR-267* :1-137, 1981.
- [12] Pople, H.E., Jr. Heuristic methods for imposing structure on ill-structured problems: the structuring of medical diagnosis. in: *P Szolovits (Ed.) Artificial intelligence in medicine. AAAS Symposia Series*. Westview Press, Boulder, Col: 119-185, 1982.
- [13] Soltan R.A., Rashad M.Z., El-Desouky B. Diagnosis of Some Diseases in Medicine via computerized Experts System. *International Journal of Computer Science & InformationTechnology (IJCSIT)* - 5(5): 79-90, 2013.
- [14] Miller RA, McNeil MA, Challinor SM, et al: The Internist-1/Quick Medical Reference project-Status report, In *Medical informatics [Special Issue]. West J Med* - 145: 816-822, 1986.
- [15] Клещев А.С., Шалфеева Е.А. Онтология задач интеллектуальной деятельности. *Онтология проектирования* –2(16): 179-205, 2015.
- [16] Denekamp Y., Peleg M. TiMeDDx - A multi-phase anchor-based diagnostic decision-support model. *Journal of Biomedical Informatics* -43: 111-124, 2010.
- [17] Технология разработки стандартов (клинических протоколов). Проект НИИ общественного здоровья и управления здравоохранением, под рук. Проф. Воробьева П.А. http://www.spruce.ru/sanatori/standard/technology/technology_5.html
- [18] Клещев А.С., Черняховская М.Ю., Москаленко Ф.М. Модель онтологии предметной области «Медицинская диагностика». Часть 1. Неформальное описание и определение базовых терминов. *Журнал НТИ. Серия 2.* - 12, 2005.