

РЕПЛИКАЦИЯ И ВАЛИДАЦИЯ В РАСПРЕДЕЛЕННОЙ СИСТЕМЕ ПОДДЕРЖКИ ВРАЧЕБНЫХ РЕШЕНИЙ¹

Л.Е. Карпов , А.Н. Томилин, В.Н. Юдин

Система «Спутник Врача», разрабатываемая в Институте системного программирования РАН, предназначена для информационной поддержки врачебных решений на основе современных информационных технологий: теории принятия решений, методов вывода по прецедентам и распознавания образов. Она призвана помочь врачу аккумулировать собственный опыт путем накопления и интерпретации данных о пациентах и их состоянии в виде прецедентов и, если необходимо, использовать опыт, накопленный другими врачами в подобных системах.

Вывод на основе прецедентов представляет собой метод принятия решений, моделирующий человеческие рассуждения. Метод использует знания о предыдущих ситуациях или случаях (прецедентах), которыми могут быть встречавшиеся ранее проблемы или типичные случаи, а также принятые тогда решения. При рассмотрении новой проблемы (текущего случая) в качестве аналога находится похожий прецедент. Решение, выбранное из аналога, либо используется прямо, либо адаптируется к текущему случаю. После того, как текущий случай будет решен, он вносится в базу прецедентов вместе со своим решением для его возможного последующего использования.

Изначально наполняемая смоделированными типовыми схемами диагностики и лечения, а в дальнейшем пополняемая случаями из врачебной практики, накопленная совокупность прецедентов образует так называемую «базу прецедентов». Система, построенная по такому принципу, является самообучаемой: чем больше прецедентов содержится в базе, тем больше спектр их возможных значений, тем выше вероятность найти «наиболее подходящий» прецедент, следовательно, выше качество принимаемого решения.

Для реализации системы выбрано двухуровневое решение в виде Оболочки и Анализатора (рис. 1). В совокупности эти уровни представляют собой презентационный и прикладной слои программного обеспечения, предназначенного для системной поддержки выполняемых работ.

Оболочка (презентационный слой) – уровень системы, реализующий интерфейс с конечным пользователем на уровне лица, принимающего решения. С помощью оболочки организуется доступ врача к данным в терминах, с которыми тот привык работать: *пациент, показатель, заболевание, лечение, и исход*. При обращении к Анализатору (прикладному слою) эти термины преобразуются в понятия *класс, объект, признак*. Анализатор выполняет основные операции анализа данных: классификацию, оценку, прогнозирование и выявление зависимостей в данных.

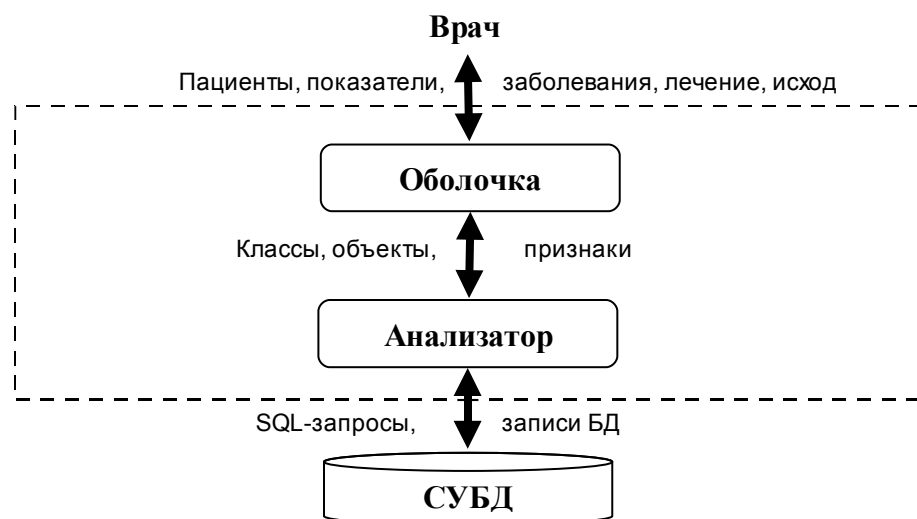


Рис. 1. Двухуровневая архитектура системы поддержки врачебных решений.

Прецеденты в Оболочке – это реальные или смоделированные случаи заболеваний. Согласно [1], всякий прецедент состоит из

1. описания некоторого случая, или проблемы;

¹ Работа выполнена при поддержке Российского Фонда Фундаментальных Исследований, проекты № 06-07-89098-а и № 06-01-00503-а.

2. описания решения проблемы (действий, предпринимаемых в этом случае);
3. результата применения решения (исход).

При выборе лечения описание случая – это симптомы больного, результаты исследований, анамнез и т. д., в целом обозначаемые как *набор признаков* больного. Признаки могут быть числовыми, логическими, даже описательными. Решение – это примененное к больному на данном шаге лечение в виде выполненных процедур. Исход, как состояние после примененного лечения, – это обратная связь, позволяющая анализировать решение. Кроме набора признаков, такой случай может содержать обширные текстовые комментарии.

Прецеденты хранятся на уровне Анализатора в виде троек. Проблема выбора наиболее подходящего прецедента является одной из самых важных в подобных системах. Искомая «мера близости», отражающая сходство текущего объекта и выбранного прецедента, определяется сравнением их признаков. Разбиение базы прецедентов на классы – один из способов ускорить поиск: прецеденты, принадлежащие одному классу, по определению, являются схожими. Классы могут изначально строиться с помощью экспертного знания, когда признаки заболеваний и границы допустимых значений этих признаков задаются экспертом на основе его теоретических знаний и имеющегося у него опыта врачебной практики, либо на основе специально разрабатываемой обучающей выборки.

Понятие *класс* в Анализаторе соответствует понятию *заболевание* в Оболочке. При рассмотрении текущего случая соответствующая ему точка сравнивается с расположением классов в проекции на пространство его признаков. Однако классы – одноуровневые понятия, поэтому на практике не всегда удается четко разграничить классы, в которые попадает текущий случай. Одной из наиболее частых причин этого является недостаток информации в описании случая.

В медицине каждое заболевание характеризуется своим набором «существенных» показателей. Разные наборы показателей могут быть не только у разных заболеваний, но и в разных случаях одного и того же заболевания. И, наконец, текущий случай может иметь набор показателей, не совпадающий с наборами показателей заболеваний, введенных в систему.

С одной стороны, ряд признаков, которыми обладает некоторый случай, может не входить в общее признаковое пространство имеющихся классов. Они считаются несущественными для данного случая и в дальнейшем не рассматриваются. С другой стороны, некоторые другие признаки, являющиеся существенными по отношению к некоторым из имеющихся классов, у текущего случая по разным причинам могут отсутствовать. Вследствие этого, неполностью описанный случай может попасть в пересечение классов, другими словами, оцениваться неоднозначно.

Отношения между случаем и классами проявляются в проекциях классов на пространство признаков случая. Случай может попасть в проекцию класса, к которому он не принадлежит, только потому, что у него не хватает признака, который дифференцировал бы его от этого класса. Классы нужно дифференцировать, добавляя значения недостающих признаков для текущего случая. В медицине подобная задача носит название «дифференциальная диагностика». На практике добавление может быть затруднено из-за дефицита средств, времени и/или оборудования.

Прецеденты можно разделить на группы, в зависимости от сложности пересечения классов, в которое попадает случай. Те прецеденты, что находятся в одной с текущим случаем внутренней области пересечения, относящейся сразу к нескольким классам, естественно считать более близкими к нему, чем те, которые находятся с ним только в одном классе (рис. 2).

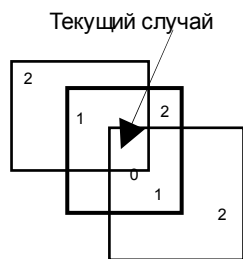


Рис. 2. Степени близости прецедентов.

При оценке набора признаков текущего случая врач не может быть уверен в том, что этот набор – достаточный, и что случай соответствует профилю врача. В медицине существует процедура «консилиума», при проведении которого набор признаков, наблюдаемых у больного, оценивается несколькими врачами, возможно, с разной специализацией. Использование информации, накопленной в нескольких системах «Спутник Врача», представляет собой определенный аналог проведения консилиума.

В Оболочке представлена иерархическая классификация типов заболеваний. В принципе, весь набор известных заболеваний представим в виде дерева. Врач работает с некоторым кустом этого дерева, соответствующим своему профилю. Например, абдоминальный хирург специализируется на заболеваниях в области живота (язвы желудка и двенадцатиперстной кишки, холецистит, панкреатит и т. д.). Накапливаемая им база содержит прецеденты этого ряда заболеваний, но сюда могут попасть прецеденты с похожими симптомами, представляющие другие заболевания, с которыми врач встретился по ходу своей практики.

В каждой отдельной системе, как правило, имеется своя структурная иерархия заболеваний (рис. 3). Это связано, во-первых, с разной специализацией врачей, во-вторых, с тем, что врач при заполнении своей базы прецедентов использует личный, во многом уникальный опыт.

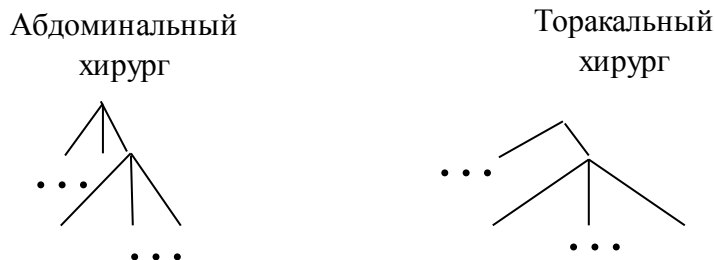


Рис. 3. Структурные иерархии заболеваний, строящиеся врачами разных специальностей.

Промежуточные узлы дерева – это группы заболеваний, листья дерева представляют собой конкретные виды заболеваний.

Однако врачи разного профиля могут иметь и пересекающиеся фрагменты структур. Каждый из них в своей практике встречается со случаями, которые, в первом предположении, были отнесены к его профилю, но дальнейшие исследования впоследствии показали, что это другое заболевание. Тем не менее, такие случаи заносятся в систему, так как они обогащают опыт врача.

Объединение нескольких баз прецедентов должно проводиться в совместном, минимально объемлющем все базы, пространстве признаков. Текущий случай оценивается на совокупности прецедентов и пересекающихся классов от нескольких систем на одном экране в основной системе (рис. 4).

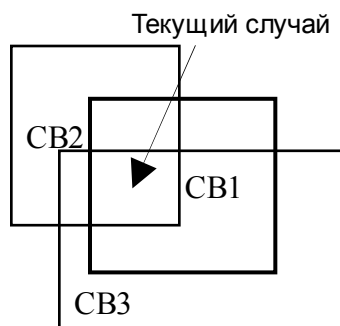


Рис. 4. Оценка текущего случая в объединенной базе прецедентов.

Такое объединение позволяет проводить оценку текущего случая, когда в объединенной системе оказываются представленными больше заболеваний, чем в каждой из составляющих ее отдельных систем. Объединение систем при рассмотрении сложного случая позволяет существенно повысить качество принимаемого решения так же, как и в случае консилиума врачей.

Можно представить два варианта использования чужого опыта в виде накопленных баз прецедентов. Первый вариант – посылка запроса в удаленную систему (рис. 5) и формирование совокупности прецедентов и пересекающихся классов из нескольких систем на одном экране в основной системе. Параметры запроса – совокупность признаков текущего случая. В ответ основная система получает совокупность схожих с текущим случаем прецедентов и описаний их классов из удаленной системы. Результирующая картина получается объединением прецедентов и классов основной и удаленной системы.



Рис. 5. Оценка текущего случая с помощью запроса к удаленной системе.

Этот случай можно трактовать как доступ к сервису удаленного компьютера. Описания классов удаленной системы воспринимаются так, как они есть. Прецеденты удаленной системы не заносятся в основную систему, а лишь используются для рассмотрения конкретного случая. Возникающая ситуация аналогична той, которая возникает при работе консилиума, когда мнение другого врача может помочь в оценке картины заболевания, но какая-либо аргументация в пользу этого мнения не требуется. При рассмотрении следующих случаев полученная информация более участвовать не будет. Работа по первому варианту имеет смысл, когда информация из параллельной системы может оказать дополнительную помощь в диагностике, но не относится прямо к области деятельности основного врача, ведущего работу с базовой системой.

Если же зафиксированный случай оказался более интересным, и врач предполагает, что в дальнейшем он будет сталкиваться с подобными случаями повторно, ему будет полезнее импортировать всю или определенную часть прецедентов удаленной системы в свою базу прецедентов. Тогда он решает позаимствовать из нее часть прецедентов вместе с описаниями их классов, либо некоторые классы целиком. Второй вариант, по-существу, заключается в репликации данных (рис. 6). В этом случае осуществляется доступ не к сервису, а к ресурсам удаленного компьютера. Прецеденты и классы из удаленной системы заносятся в основную систему, что приводит к необходимости полностью реорганизовать базу прецедентов основной системы для переопределения границ классов с учетом возможных коллизий. Реорганизация может привести к выявлению новых классов, ранее отсутствовавших в реплицируемой базе.

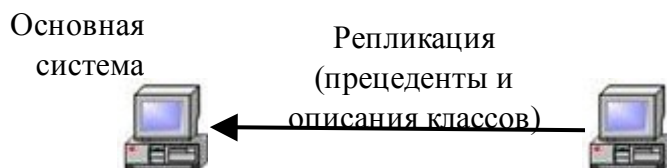


Рис. 6. Импорт информации из удаленной системы.

Оба рассмотренных варианта имеют очень важную модификацию, существенным образом, влияющую на процесс взаимодействия систем между собой. Эта модификация процесса распространения информации между системами состоит в том, что в качестве одного из источников дополнительных данных может выбираться крупный информационный центр коллективного доступа, который содержит выверенные данные, поступающие из локальных систем различных практикующих врачей (рис. 7).

Заполнение коллективной базы прецедентов, которая используется врачами, имеющими различный опыт, не может осуществляться путем простой репликации данных. В данном случае репликация должна в обязательном порядке сопровождаться также валидацией описаний прецедентов, то есть синхронизацией терминологии, уточнением численных и логических показателей, отсеиванием сомнительных и откровенно ошибочных записей.

Основная проблема в рассмотренных случаях – различия в обозначении заболеваний (классов) и показателей (признаков). В разных базах прецедентов, заполненных разными врачами, могут использоваться разные наименования одного и того же признака, разные написания одного наименования (например, двойной пробел, тире-минус и т. д.). Та же проблема может возникать и с заболеваниями. Считать, что врачи, разного профиля сталкиваются только с разными заболеваниями, неправомерно. Как допустим обмен информацией между врачами одного профиля, также допустим подобный обмен и между врачами разных профилей. Кроме того, базы прецедентов, по-

строенные врачами одного профиля, могут иметь прецеденты с описаниями нехарактерных для них заболеваний, являющихся при этом типичными для других направлений врачебной практики.

Рис. 7. Совместное использование локальных систем и информационных центров коллективного доступа.



Та же проблема, хотя и не столь критичная, возникает из-за того, что встречается много производителей одного и того же препарата с разными названиями (перечень использованных препаратов - это часть прецедента, относящаяся к *решению*, которая носит, в основном описательный характер). Эта проблема не возникнет, если врач относится аккуратно к внесению прецедента в базу, указывая в решении не только фирменное, но и международное принятое обозначение, которое всегда присутствует в описании препарата.

Для осуществления любых видов взаимодействия составных частей (центральной части коллективного пользования и локальных систем, находящихся в персональном пользовании отдельных врачей) совокупной распределенной системы наиболее важной является стандартизация терминологии, наименования заболеваний и признаков (например, через словарь признаков), а также методов их описания. В отдельных случаях система должна быть готова работать в условиях наличия различий наименований заболеваний.

Перспективы использования описанной технологии находятся в создании систем поддержки принятия решений, удаленный доступ к которым может получить врач любой специальности, сталкивающийся с нозологическими формами, отличными от привычных ему форм, и, таким образом, оптимизировать диагностику, лечение и реабилитацию наблюдаемых им больных.

Чтобы обеспечить возможность такого удаленного доступа, как в локальных, так и в глобальных сетях (в том числе в сети Интернет), при разработке программного обеспечения предлагается использовать совокупность стандартов, на которых базируются современные сетевые службы. Изучение тенденций развития сетевых служб показывает, что их разработчики все чаще стремятся использовать единые подходы, независимо от того, предполагается ли работать только в локальных сетях, или возможен выход и в глобальные сети (корпоративные или общедоступные).

Дальнейшая работа группы разработчиков направлена на использование системы удаленного доступа через Интернет к информационным центрам коллективного доступа, в которых аккумулирован большой объем данных (применительно к медицине – это крупные медицинские центры). Это позволит широкому кругу исследователей пользоваться накопленными там данными, используя содержащиеся в них реальные объекты и ситуации как прецеденты.

ЛИТЕРАТУРА:

1. Klaus-Dieter Althof, Eric Auriol, Ralph Barlette, and Michel Manago. A Review of Industrial Case-Based Reasoning Tools. AI Intelligence, 1995.

2. Л. Е. Карпов, В. Н. Юдин. Методы добычи данных при построении локальной метрики в системах вывода по прецедентам. Институт Системного Программирования РАН, Препринт, 2006.
3. Л. Е. Карпов, В. Н. Юдин. "Адаптивное управление по прецедентам, основанное на классификации состояний управляемых объектов", М., Труды Института Системного Программирования РАН, 2007, т. 13, ч. 2, стр. 37-57.
4. Л. Е. Карпов, В. Н. Юдин. "Интеграция методов добычи данных и вывода по прецедентам в медицинской диагностике и выборе лечения". Сборник докладов 13-й Всероссийской конференции "Математические методы распознавания образов (ММО-13)", октябрь, 2007, стр. 589-591.
5. Л. Е. Карпов, В. Н. Юдин. "Система поддержки принятия решений для практикующих врачей". Ежегодная техническая конференция «Корпоративные базы данных-2008», <http://www.citforum.ru/seminars/cbd2008/>
6. Юдин В. Н., Карпов Л. Е., Ватазин А. В. "Процесс лечения как адаптивное управление человеческим организмом в программной системе «Спутник Врача»", М., Альманах клинической медицины, 2008, т.17, часть 1, стр. 262-265.
7. Юдин В. Н., Карпов Л. Е., Ватазин А. В. "Методы интеллектуального анализа данных и вывода по прецедентам в программной системе поддержки врачебных решений", М., Альманах клинической медицины, 2008, т.17, часть 1, стр. 266-269.