

# Обмен данными в распределенной системе поддержки решений<sup>1</sup>

Карпов Л. Е., Юдин В. Н.

**Аннотация.** Обмен данными – расширение возможностей локальной системы поддержки решений с целью привлечения опыта, накопленного другими пользователями в подобных системах. Рассматриваются два варианта обмена в сети, где присутствуют несколько подобных систем: виртуальная интеграция (аналог консилиума) и консолидация (импорт знаний).

## 1. Введение

Система поддержки принятия решений разрабатываемая в Институте системного программирования РАН, предназначена для помощи пользователю-эксперту в аккумуляции его опыта путем накопления и интерпретации его знаний в виде *прецедентов* (случаев) и информационной поддержки принятия решений в различных областях интеллектуальной деятельности на основе современных информационных технологий: теории принятия решений, вывода по прецедентам и распознавания образов.

Изначально наполняемая смоделированными типовыми схемами принятия решений, а в дальнейшем пополняемая случаями из реальной практики, накопленная совокупность прецедентов не является простым набором слабо связанных между собой случаев. Напротив, эта совокупность структурируется и классифицируется функционирующей системой, образуя так называемую *базу прецедентов*. Система, построенная по такому принципу, является самообучаемой: чем больше прецедентов содержится в базе, тем характернее спектр возможных значений их параметров, тем выше вероятность найти *наиболее подходящий* прецедент, тем, следовательно, выше качество принимаемого решения.

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 в связи с ними решения.

Традиционно прецедент рассматривается, как набор, состоящий из описания некоторого случая или проблемы, описания решения проблемы (действий, предпринимаемых в этом случае) и результата применения решения (исход). При рассмотрении новой проблемы (текущего случая) в системе делается попытка найти похожий прецедент, который в дальнейшем надлежит использовать в качестве аналога. Решение проблемы, выбранное из аналога, частью которого оно является, либо используется прямо и непосредственно, либо адаптируется к текущему случаю. После того, как текущий случай будет решен (то есть для него будет сформировано решение и изучен его результат), этот случай вносится в базу прецедентов вместе со своим решением и исходом для его возможного последующего использования.

В основе подходов к отбору прецедентов лежит оценка схожести прецедента и текущего случая. Обычно для определения этой схожести в пространстве признаков вводится специальная метрика, для вычисления которой в этом пространстве определяются точки, соответствующие текущему случаю и его исследуемым аналогам. На основе вычисленных значений метрики находится ближайшая к текущему случаю точка, представляющая прецедент.

В реальной жизни, когда объекты описываются разнородными признаками, в том числе и логическими, ввести метрику, которая обладает строго определенными свойствами, не всегда удается. В этих случаях вместо метрики используется так называемая *мера близости*. Один из способов применения меры близости использует разбиение базы прецедентов на классы, то есть группы "похожих" случаев (случаи, принадлежащие одному классу, по такому определению, являются схожими). Разбиение базы прецедентов на отдельные классы может проводиться по-разному. Например, такое разбиение можно проводить с помощью экспертного знания, когда признаки заболеваний и границы допустимых значений этих признаков задаются экспертом-врачом на основе его теоретических знаний и имеющегося у него опыта врачебной практики, можно также использовать специально разработанные обучающие выборки случаев.

База прецедентов системы поддержки принятия решений включает в себя:

- описание объемлющего признакового пространства для случаев, хранящихся в базе, в частности, типы и границы признаков объектов,
- описания случаев, рассматриваемых как прецеденты (признаки случая, решение, исход),
- описания классов случаев (перечень признаков класса, границы признаков).

Основные функции системы поддержки – оценка (распознавание) текущего случая, представленного совокупностью признаков, и поиск аналогов случая.

<sup>1</sup> Поддержка работы осуществляется Российским Фондом Фундаментальных Исследований по проектам № 09-01-00351-а и № 09-07-00191-а.

Текущий случай сравнивается с описаниями классов в базе прецедентов, что позволяет отнести его к тому или иному классу (задача отнесения отдельного объекта к одному из классов называют распознаванием объекта). В разрабатываемой программной системе реализуется метод отбора наиболее подходящих прецедентов, базирующийся на оценке близости, смысл которой описан далее в этой работе.

Каждый класс в базе прецедентов представлен в пространстве признаков многомерным параллелепипедом, минимально объемлющим прецеденты этого класса. При поиске подходящего класса значения признаков текущего случая сравниваются с проекциями границ классов на пространство этих признаков. В общем случае при таком сравнении выделяется сразу целый ряд классов, в которые попадает исследуемый случай, для работы выбираются те прецеденты, которые находятся в любом из них.

В описании случая может присутствовать большое количество признаков, на практике их может быть до нескольких десятков, поэтому четко разграничить классы, в которые попадает случай, удастся не всегда. Одной из наиболее частых причин этого является недостаток информации в его описании. В подобных случаях среди признаков выделяют *существенные*. Может также возникнуть ситуация, в которой текущий случай имеет набор признаков, не пересекающийся с наборами признаков, ранее введенными в систему. Иногда такая ситуация возникает вследствие того, что некоторые несущественные для данного случая признаки были сознательно отброшены и не включены в рассмотрение. Однако некоторые другие признаки, являющиеся существенными по отношению к некоторым из имеющихся классов, у текущего случая по разным причинам могут также отсутствовать. Вследствие этого, не полностью описанный случай может попасть в пересечение классов (то есть сразу в несколько классов одновременно), другими словами, оцениваться неоднозначно только потому, что у него не хватает признака, который дифференцировал бы его от других классов. В зависимости от сложности пересечения классов, в которое попал исследуемый случай, прецеденты делятся на группы. Те прецеденты, что находятся во внутренней области пересечения, естественно считать наиболее близкими к текущему случаю, их в проводящемся анализе ситуации необходимо рассматривать в первую очередь, обращаясь к другим (временно отложенным прецедентам) только в случаях крайней необходимости.

Важным свойством создаваемой системы поддержки принятия решений является возможность использования опыта нескольких экспертов одновременно. С технической стороны это приводит к возможности создания распределенных конфигураций системы, работающих в локальных и/или глобальных общедоступных или специализированных информационных сетях. В то же время сохраняется возможность создавать изолированные установки, работающие в локальном окружении, без каких-либо контактов с другими аналогичными установками системы. В системе, предназначенной

для персонального использования на локальном компьютере, каждый пользователь хранит описания встретившихся именно ему прецедентов, формирует собственную базу прецедентов, не имея доступа к опыту других пользователей. Однако естественным расширением возможностей системы было бы привлечение опыта, накопленного другими пользователями в подобных системах. Этот опыт материализован в данных, хранящихся в других установках системы. Обмен данными между аналогичными системами, по сути, представляет собой обмен накопленными знаниями.

Под распределенной системой поддержки решений будем понимать совокупность локальных систем, между которыми имеется возможность обмена данными. Метод пересылки данных при нашем рассмотрении архитектуры распределенной системы значения не имеет. Для передачи параметров запроса могут быть использованы разные технические методы: от обмена файлами на физическом носителе – до возможностей информационных сетей.

Обмен может осуществляться двумя способами. Первый вариант – запрос из базовой системы в удаленный компьютер. Параметры запроса – показатели текущего случая (Рис. 1): имя системы и описание случая, то есть значения его признаков.



Рис. 1. Оценка текущего случая с помощью запроса к удаленному компьютеру.

В ответ на запрос система с удаленного компьютера сообщает базовой системе схожие прецеденты вместе с описаниями их классов:

- Имя системы (необязательный параметр сообщения)
- Группы описаний классов, каждая из которых содержит
  - имя класса
  - границы класса по признакам
- Группы описания прецедентов, каждая из которых содержит
  - имя случая
  - признаки случая

Пользователь не обязан знать, на каком компьютере данные хранятся физически. В запросе, адресованном конкретной системе в сети, где присутствуют несколько подобных систем, имеется возможность указывать имя системы, но если имя не указано, это означает, что запрос выдается ко всей сети в целом. Системы, где выявлены описания схожих случаев, выдают ответ на запрос в указанном формате. Чтобы базовая система не ждала ответа от сети бесконечно долго, остальные выдают пустой ответ.

В базовой системе формируется картина из прецедентов и классов, полученная в результате объединения всей имеющейся и полученной в ответ на запрос информации (Рис. 2), так, как если бы все данные хранились в одной системе.

Такой способ обмена будем называть *виртуальной интеграцией*, так как реального пополнения базы прецедентов базовой системы при рассмотрении конкретного случая не происходит. Виртуальная интеграция позволяет проводить на практике оценку текущего случая, принимая во внимание опыт, накопленный в каждой из локальных систем. В классической медицине, например, аналогом такой интеграции является консилиум врачей, когда мнение другого врача может помочь в оценке картины заболевания. Объединение данных из нескольких систем при рассмотрении сложного случая позволяет существенно повысить качество принимаемого решения так же, как и в случае консилиума.

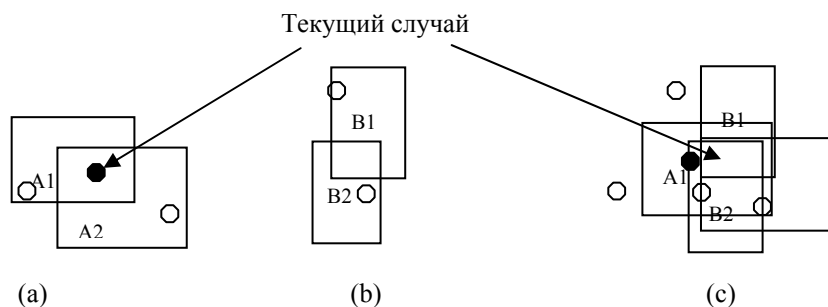


Рис. 2. Оценка текущего случая методом виртуальной интеграции

- a) в базовой системе (A) до интеграции, два прецедента из классов A1 и A2
- b) в удаленном компьютере (B), два прецедента из классов B1 и B2
- c) в базовой системе (A) после интеграции.

В частном случае, в базовой системе вообще могут отсутствовать прецеденты и классы, необходимые для оценки текущего случая. Тогда виртуальная картина будет собираться из того, что получено по запросу из удаленных компьютеров.

Реально картина на экране компьютера не выглядит так просто, как на рис. 2. Многомерный случай размерностью больше двух невозможно достаточно наглядно представить двумерной проекцией. Наиболее информативный вариант – представление в виде так называемой реляционной модели, в виде нескольких взаимосвязанных таблиц. Покажем это на примере системы поддержки врачебных решений в диагностике и выборе лечения «Спутник Врача», разрабатываемой в Институте системного программирования РАН в рамках текущего проекта.

На Рис. 3 показаны прецеденты, найденные системой для смоделированного случая «симптомы острого живота» (левый верхний угол). Источником двух первых прецедентов является базовая система (которую заполнял абдоминальный хирург, специализирующийся на лечении органов брюшной полости), третьего – удаленный компьютер (заполнял пульмонолог – специалист, занимающийся лечением заболеваний лёгких и дыхательных путей).

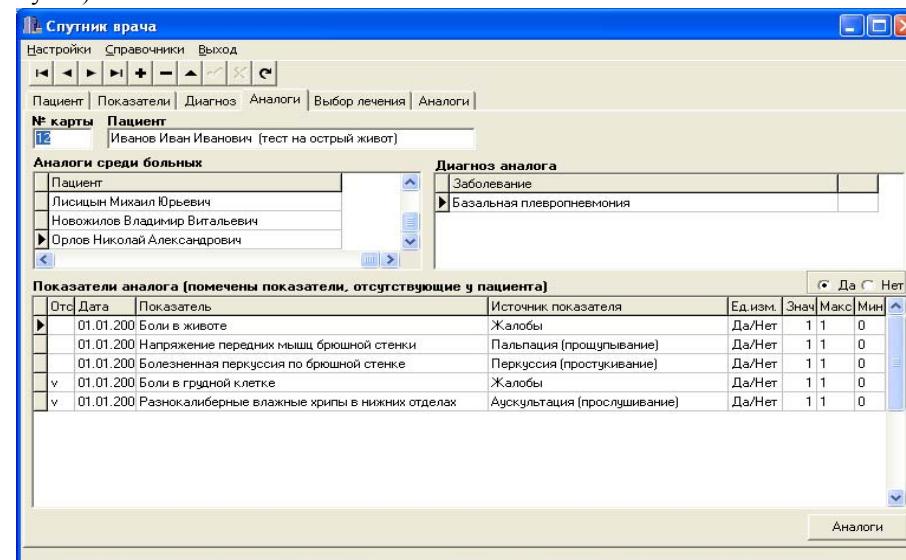


Рис. 3. «Спутник врача». Прецеденты с симптомами «острого живота».

Виртуальная интеграция имеет смысл, когда информация из удаленной системы может оказать дополнительную помощь в диагностике, но не относится прямо к области деятельности пользователя, ведущего работу с базовой системой. При рассмотрении последующих случаев полученная информация более участвовать не будет (виртуальная интеграция всегда должна проводиться заново при рассмотрении каждого нового случая).

Если же зафиксированный случай оказался более интересным, и пользователь базовой системы 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. Возникает второй вариант обмена – *консолидация знаний*. В этом случае осуществляется доступ не к сервису, а к данным удаленного компьютера (рис. 4).

Консолидация – это не просто импорт части случаев из одной базы прецедентов в другую, но и сопутствующая ему реорганизация базы прецедентов. Импорт случаев из базы может выполняться как в автоматизированном, так и в ручном режиме. Прецеденты и классы из удаленной системы копируются в базовую систему, что приводит к необходимости переопределения границ классов с учетом возможных коллизий. Реорганизация может привести к выявлению новых классов, ранее отсутствовавших в базовой системе.

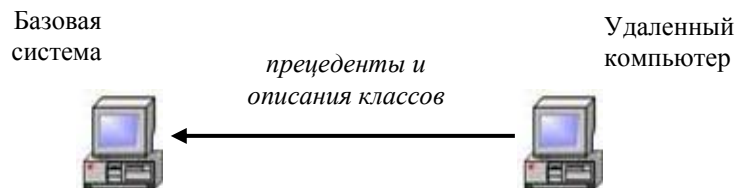


Рис. 4. Импорт прецедентов из удаленного компьютера.

Пользователь базовой системы может контролировать процесс импорта. Сначала на экране системы высвечивается картина, отражающая содержимое, полученное по запросу от удаленного компьютера (Рис. 2б). Оператор помечает прецеденты, которые нужно импортировать, затем переключается в интегрированный режим для просмотра ожидаемого результата (Рис. 2с). При необходимости цикл повторяется многократно.

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 (Рис. 5).

Заполнение коллективной базы прецедентов не может осуществляться путем простого импорта данных. В данном случае импорт должен в обязательном порядке сопровождаться *валидацией* описаний случаев, то есть

синхронизацией терминологии, уточнением численных и логических показателей, отсеиванием сомнительных и откровенно ошибочных данных.

Чтобы обеспечить возможность удаленного доступа пользователей к системе и облегчить его технически, 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).

На примерах взаимодействия нескольких установок разрабатываемой системы поддержки принятия решений можно проследить аналогию с организацией распределенных баз данных и связанным с ними понятием *репликации*. Общим для этих понятий является то, что система поддержки принятия решений должна позволять управлять распределенными данными таким образом, чтобы эта распределенность была прозрачна для пользователей. Прозрачность доступа означает, что пользователи осуществляют доступ к распределенным данным точно так же, как если бы они хранились централизованно.



Рис. 5. Совместное использование локальных систем и информационных центров коллективного доступа.

Отличия рассматриваемых понятий – значительные. Распределенные системы поддержки принятия решений – самодостаточные системы, созданные для использования независимыми пользователями. В этих системах отсутствует разделение приложений и базы данных по узлам приема запросов и узлам данных. Словари данных устроены однотипно, но создаются они независимо:

классы, одинаковые по сущности, в разных системах могут иметь разное обозначение и несовпадающие границы.

Если репликация – это процесс приведения данных электронных таблиц двух баз данных в идентичное состояние, то обмен данными в системах поддержки принятия решений лучше назвать *управляемым копированием*. Если репликация в распределенных базах данных может быть однонаправленной или мультинеправленной, то здесь обмен всегда однонаправленный. По старшинству локальные системы никак не выделяются: все взаимодействующие системы считаются равноправными, обмен может происходить между любой парой из них.

Для распределенных баз данных существует проблема возобновления работы процесса репликации при потере связи и последующем ее восстановлении. Здесь же данные обладают свойством идемпотентности: обмен данными происходит одним файлом, при обрыве связи передача просто повторяется.

Репликация в базах данных может быть синхронной или асинхронной. Процесс обмена в системах принятия решений всегда синхронный.

Проблема справочников и генерации идентификаторов не столь важна в системах принятия решений. Конфликты тоже возникают, но в основном – в описаниях классов и случаях отнесения к тому или иному классу. Различия в обозначении и границах классов не имеют большого значения. Наоборот, они позволяют пользователю сопоставить свой и чужой опыт при оценке текущего случая. Эти конфликты могут разрешаться либо на уровне алгоритма обмена, использующего правила административного старшинства (например, базовая система считается приоритетной), либо, руководствуясь наиболее поздним по времени событием становления класса, либо на уровне вмешательства ответственного пользователя - администратора базовой системы. Но даже при избытке первых двух факторов, все равно невозможно разрешить все конфликты без участия администратора.

## Литература

- [1] Карпов Л. Е., Юдин В. Н. Методы добычи данных при построении локальной метрики в системах вывода по прецедентам // Институт Системного Программирования РАН, Препринт, 2006.
- [2] Карпов Л. Е., Юдин В. Н. Адаптивное управление по прецедентам, основанное на классификации состояний управляемых объектов // Труды Института Системного Программирования РАН, т.13, ч.2. М., 2007, стр. 37-57.
- [3] Карпов Л. Е., Юдин В. Н. Интеграция методов добычи данных и вывода по прецедентам в медицинской диагностике и выборе лечения // Сб. докладов 13-й Всероссийской конференции «Математические методы распознавания образов (ММРО-13)», М., 2007, стр. 589-591.
- [4] Карпов Л. Е., Юдин В. Н. Система поддержки принятия решений для практикующих врачей // Ежегодная техническая конференция «Корпоративные базы данных-2008», Электронный ресурс. Режим доступа: <http://www.citforum.ru/seminars/cbd2008>

- [5] Юдин В. Н., Карпов Л. Е., Ватазин А. В. Методы интеллектуального анализа данных и вывода по прецедентам в программной системе поддержки врачебных решений // Альманах клинической медицины, т.17, часть 1. М., 2008, стр. 266-269.
- [6] Ватазин А. В., Карпов Л. Е., Юдин В. Н. Виртуальная интеграция и консолидация знаний в распределенной системе поддержки врачебных решений // Альманах клинической медицины, т.20, М., 2009, стр. 83-86.
- [7] Watson I., Marir F. Case-Based Reasoning: A Review, The Knowled. Engineer. A Review, 1994. V. 9. No.4.