

Витяев Е.Е.

Компьютерное познание¹

В работе анализируются методы интеллектуального анализа данных и их связь с процессом познания. Критикуется аппроксимационный подход к познанию предметных областей сложившийся в этом направлении. Анализируется возможность машинного познания предметных областей, основываясь на результаты теории измерений. Проведённый анализ приводит к определённой методологии познания предметных областей, которая представляет собой определённый процесс компьютерного познания (за рубежом это направление исследований называется Scientific Discovery). Приводится результирующая схема компьютерного познания.

Ключевые слова: Интеллектуальный анализ данных, Machine Learning, Knowledge Discovery in Data Bases and Data Mining, KDD&DM, Data Mining, искусственный интеллект, знание, познание.

1. Процесс познания, основанный на теории измерений

В настоящее время интенсивно развиваются такие направления как: Интеллектуальный Анализ Данных (ИАД) и Knowledge Discovery in Databases and Data Mining (KDD&DM), основанные на методах Машинного Обучения (Machine Learning), Искусственного Интеллекта (Artificial Intelligence) и Анализа данных (Data Analysis). Давно назрела необходимость проанализировать эти методы с точки зрения их связи с процессом познания. В данной работе проводится такой анализ, который приводит к определённой методологии познания предметных областей [Витяев Е.Е., 2006], а именно к компьютерному познанию (за рубежом это направление исследований называется Scientific Discovery).

В работе приводится обзор разработанной методологии и даётся множество ссылок на конкретные теоретические и практические работы, в которых подробно доказывается, поясняется и иллюстрируется эта методология. Ссылки сделаны на сайт [website Scientific Discovery], разработанный совместно с prof. B. Kovalerchuk (Central Washington University), с которого статьи скачиваются.

1. Аппроксимационный подход в ИАД. Во всех существующих методах ИАД исходное множество гипотез задается априори. В нейронных сетях это кусочно-линейные функции, в деревьях это логические решающие функции, в регрессионном анализе это линейная или нелинейная регрессия, в дискриминантном анализе – дискриминантная функция, в распознавании образов – решающее правило, методах классификации – форма кластеров. Какова в некотором смысле «истинная» зависимость – такой вопрос не рассматривается. Таким образом, в методах KDD&DM неизвестная зависимость аппроксимируется некоторым заданным априори классом функций, моделей, решающими правилами и т.д.

Процесс аппроксимации начинается с переноса способов измерения из точных наук в другие области. Рассмотрим, например, такую физическую величину как температура. Шкалы температуры в нефизических областях, например, при измерении температуры тела больного в медицине, температуры почвы в сельском хозяйстве, температуры воздуха в духовке в кулинарии и т.д., должны быть разные, хотя измеряться они могут одним и тем же прибором – термометром. Далеко не всеми понимается тот факт, что шкала – это набор операций и отношений, которые имеет смысл производить с числовыми значениями величин в данной Предметной Области (ПО). Точнее, это те операции и отношения, которые интерпретируемы в системе понятий (онтологии) соответствующей предметной области. Можно возразить, что термометр не может измерять ничего кроме температуры. Он действительно во всех случаях измеряет физическую температуру. Но зачем мы измеряем температуру? Ведь не затем чтобы согласно законам физики узнать, сколько в больном содержится тепла, и не затем, чтобы определить среднюю кинетическую энергию молекул почвы или курицы в духовке. Температура, как и любой другой прибор, нужны для получения выводов в системе понятий (онтологии) той предметной области, к которой он относится. Для больного «температурный фактор служит наиболее общим и универсальным регулятором скорости химических реакций и активности ферментов, с повышением температуры в известной мере ускоряются и обменные процессы» [0]. Для почв температура должна интерпретироваться в системе понятий физиологии растений и деятельности микроорганизмов и

¹ Работа поддержана грантом РФФИ 08-07-00272-а; интеграционными проектами СО РАН №1, 115, а также Госконтрактом 2007-4-1.4-00-04 и Советом по грантам Президента РФ и государственной поддержке ведущих научных школ (проект НШ-335.2008.1).

т.д. В данном случае *физическая величина* температуры является *косвенным измерением* некоторой другой величины, интерпретируемой в системе понятий предметной области, которую мы и хотим измерить. Физическая температура больного – есть косвенное измерение медицинской величины – уровня обмена веществ, температура почвы измеряет состояние биохимических процессов в растениях и микроорганизмах, температура воздуха в духовке измеряет течение процесса свертывания белка и т.д. Какие отношения и операции над числовыми значениями температуры имеют смысл для всех этих величин, определяется уже этими интерпретациями. Поэтому числовые значения величин нельзя просто так переносить из одной области знаний в другую. После такого переноса необходимо заново определять шкалу. Например, для температуры больного интерпретируемы выделенные значения 36.7, 42. и отношение линейного порядка $<$, поэтому это будет шкала порядка с выделенными значениями.

2. Построение «истинных» величин, законов и моделей. Для того чтобы детальнее разобраться с такими понятиями как числовые значения величин, их интерпретируемость, осмысленность математических операций с величинами, «истинная» зависимость и т.д., необходимо обратиться к *теории измерений* [Пфанцгль И., 1976; Krantz D.H. et al, 1971]. Теория измерений основана на принципе – *свойства определяются отношениями*. Из теории измерений следует, что *числовые значения величин и функциональные выражения для законов являются лишь удобным и математически хорошо разработанным способом числового кодирования элементов эмпирических систем*. Эмпирическая (многосортная) система $\mathfrak{S} = \langle A, \Omega \rangle$ это множество (идеализированных) объектов A , с заданным на нём множеством Ω отношений и операций, интерпретируемых в онтологии ПО. Число, например, 5 само по себе смысла не имеет, оно приобретает смысл лишь при его интерпретации в некоторой эмпирической системе, например, если мы говорим 5 метров, 5 баллов, 5 деталей и т.д. Интерпретация чисел, в частности, определяет какие математические действия с ними осмысленно можно проводить, чтобы не получать бессмысленных результатов типа 1.5 дровосека, 1м. + 1кг., и т.д.

Таким образом, исследуемую предметную область можно представить эмпирической системой $\mathfrak{S} = \langle A, \Omega \rangle$, где A – объекты ПО, а Ω – множество отношений и операций, интерпретируемых в онтологии ПО. Онтология ПО определяет «взгляд», «точку зрения», с которой рассматриваются объекты предметной области. Понятия онтологии могут быть выражены одноместными предикатами и также могут входить в Ω . Такой *качественный (семантический) уровень рассмотрения ПО с необходимостью возникает из того факта, что интерпретировать человек может только качественно – в онтологии предметной области*. Поэтому, при интерпретации количественных значений величин, моделей, функций и т.д. он *интерпретирует их качественно – в системе понятий предметной области (онтологии)*. Отсюда возникает *качественная эмпирическая теория* (рис. 1).

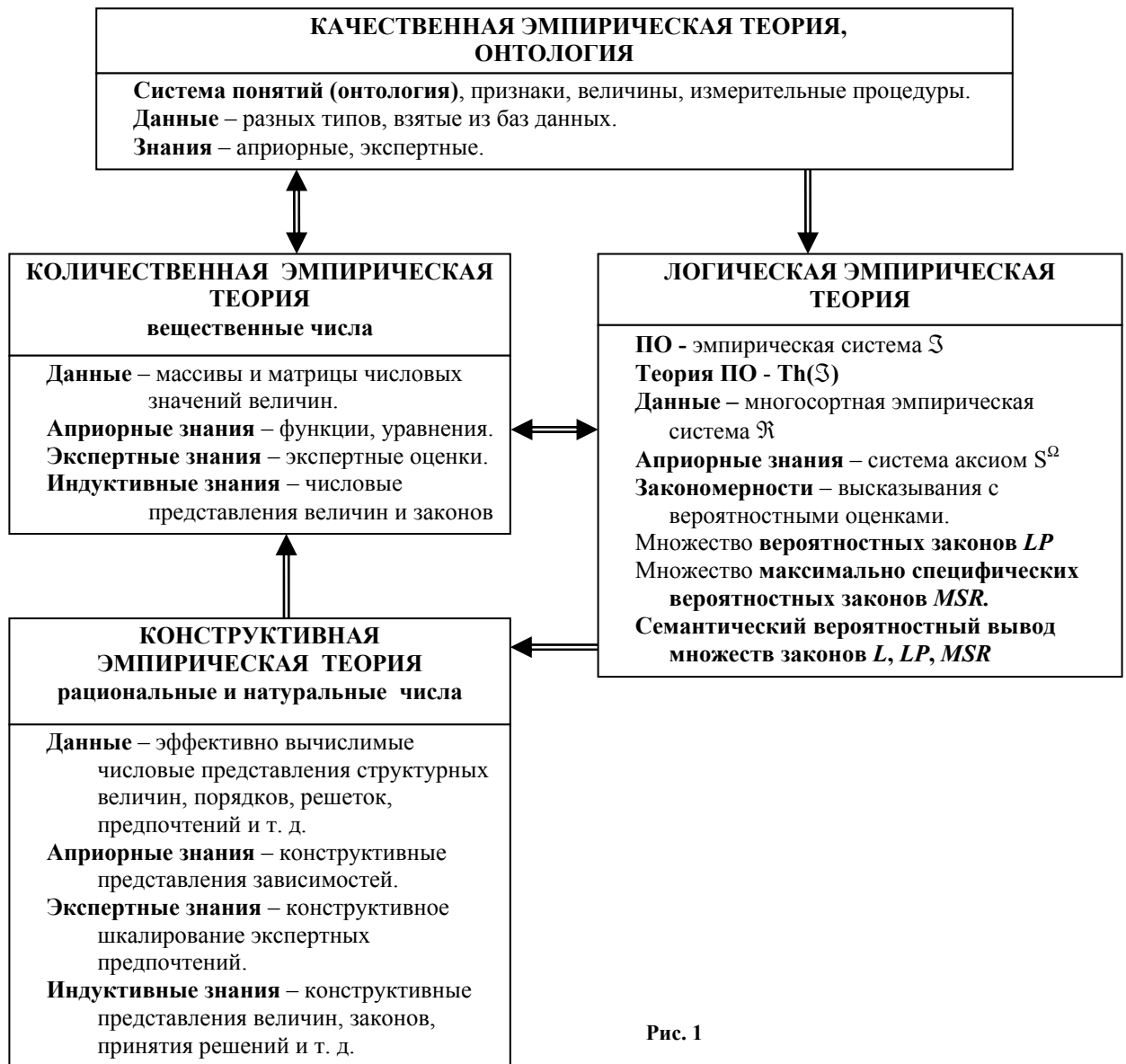


Рис. 1

В отличие от аппроксимационного подхода, в теории измерений определяются в некотором смысле «истинные» величины и законы. Числовые представления величин и законов [Пфанцагль И., 1976], «истинны» в том смысле, что они интерпретируемы в системе понятий предметной области и являются лишь *числовыми кодами значений величин(ы)* соответствующей эмпирической системы.

Числовые представления законов в теории измерений «истинны» в том смысле что они:

(1) интерпретируемы в системе понятий предметной области и являются лишь числовыми кодами взаимосвязей величин эмпирической системы;

(2) получаются процедурой одновременного шкалирования закона и числовых представлений входящих в него величин. В [Пфанцагль И., 1976; Krantz D.H. et al, 1971] показано, что физические законы просты потому, что они являются результатом одновременного шкалирования всех, входящих в зависимость величин так, что взаимосвязь этих величин выражается заданной (определяемой системой аксиом) простой функциональной зависимостью.

Поскольку эмпирическая система \mathcal{S} представляет предметную область, то теория предметной области $Th(\mathcal{S})$ – это множество истинных на эмпирической системе высказываний. В дополнении к теории ПО, мы будем рассматривать также вероятностные высказывания (*закономерности*) в логике первого порядка, выполнимые на эмпирической системе \mathcal{S} . Точное определение закономерностей приведено в работах [Витяев Е.Е., 1991, 2006; E. Vityaev, B. Kovalerchuk, 2004]. Теория $Th(\mathcal{S})$ вместе с множеством закономерностей, которым удовлетворяют эмпирические системы, представляют собой *логическую эмпирическую теорию* (ЛЭТ) предметной области (рис. 1). Задача познания предметной области состоит в усилении (в логическом смысле) ЛЭТ.

Процесс обнаружения ЛЭТ описан в [Витяев Е.Е., 1991, 2006; E. Vityaev, B. Kovalerchuk, 2004].

Каждый ИАД-метод имеет свою онтологию [B. Kovalerchuk, E. Vityaev, 2008] – систему понятий в которой интерпретируются данные и гипотезы, проверяемые методом. Логический анализ онтологии некоторого ИАД-метода, даёт класс гипотез (закономерностей), которые обнаруживает этот метод. Эти закономерности можно использовать для познания ПО, этим методом.

Числовые представления величин и законов можно получить, применяя теорию измерений к ЛЭТ. Для этого надо проверить, какие системы аксиом теории измерений, описывающие величины и законы, содержатся в ЛЭТ. Если есть некоторая система аксиом теории измерений в ЛЭТ, выполнимая на эмпирической системе, то, применяя соответствующий результат теории измерений о существовании числового представления соответствующей величины или закона, можно получить их числовые представления. Полученные шкалы величин и законы дают *количественную теорию предметной области* (КЭТ) (рис. 1). Для физики этот переход продемонстрирован в [Krantz D.H. et al, 1971], где показано как можно создавать КЭТ в виде системы величин, связанных между собой фундаментальными законами.

Таким образом, проделанный анализ показывает, что задача познания предметной области разбивается на два этапа: сначала надо построить логическую эмпирическую теорию, а затем, применяя теорию измерений или теорию конструктивных моделей (см. ниже), построить количественную теорию предметной области. Такое разбиение отражает естественный процесс *перехода теории из качественного состояния*, представленного логической эмпирической теорией, *в количественное*. Теория измерений и является теорией такого перехода.

2. Задачи процесса познания ПО.

Рассмотрим подробнее, как должен осуществляться процесс познания некоторой предметной области. Для этого надо сначала задать предметную область. Задание предметной области осуществляется заданием *онтологии* (см. рис. 1), которая включает:

- систему понятий ПО, в которой формулируется и интерпретируется эмпирическая теория;
- свойства, признаки, величины, данные, измерительные процедуры, интерпретируемые в системе понятий.

Качественная эмпирическая теория кроме того включает:

- априорные и экспертные знания;
- знания, получаемые в процессе построения логической, количественной и конструктивной эмпирических теорий после их интерпретации в системе понятий ПО.

2.1. Построение логической эмпирической теории. Сначала надо представить предметную область эмпирической системой $\mathfrak{S} = \langle A, \Omega \rangle$. Для этого надо решить следующую задачу.

Задача 1. Определить множество Ω отношений и операций для всех свойства, признаков, величин, данных и измерительных процедуры, интерпретируемых в онтологии. Понятия онтологии также нужно представить одноместными предикатами и включить во множество Ω . Представить данные в виде (многосортной) эмпирической системы $\mathfrak{X} = \langle B, \Omega \rangle$, $B \subset A$, B – конечно.

Для решения этой задачи в [Витяев Е.Е., 1991, 2006] показано, как такие известные типы данных, как парные и множественные сравнения, матрицы упорядочений, матрицы близости и матрицы объект-признак могут быть представлены (многосортными) эмпирическими системами. Там же приведены результаты теории измерений, относящиеся к соответствующим отношениям и операциям, используя которые можно найти «истинные» числовые представления соответствующих величин.

Априорные знания качественной эмпирической теории также нужно представить в ЛЭТ системой аксиом S^{Ω} .

Экспертные знания могут быть извлечены из эксперта разными методами. Один из методов извлечения из эксперта полного и непротиворечивого множества знаний приводится в [Kovalerchuk, B., Vityaev E., Ruiz J.F., 2001; Витяев Е.Е., 2006].

После представления ПО эмпирической системой \mathfrak{S} возникает задача обнаружения логической эмпирической теории, включающей $Th(\mathfrak{S})$ и множество закономерностей.

В дальнейшем будем предполагать, что ЛЭТ представима совокупностью универсальных формул. Это ограничение, во-первых, имеет обоснование, приведенное в [Витяев Е.Е., 2006], во-вторых, есть методы его обхода и, в-третьих, можно обнаруживать также высказывания с квантором существования (см. пример в [Витяев Е.Е., 2006]).

Известно, что множество универсальных формул логически эквивалентно множеству формул вида

$$\forall x_1, \dots, x_k (A_1^{\varepsilon_1} \& \dots \& A_k^{\varepsilon_k} \Rightarrow A_0^{\varepsilon_0}), k \geq 0, \quad (1)$$

где A_0, A_1, \dots, A_k – атомарные формулы, $\varepsilon_i = 0(1)$, $i = 0, 1, \dots, k$, если атомарная формула берется без отрицания (1), с отрицанием (0).

Потому для обнаружения ЛЭТ достаточно уметь обнаруживать формулы вида (1). Экспертные и априорные знания S^{Ω} также можно преобразовать в совокупность формул вида (1). Потому в общем случае метод обнаружения ЛЭТ является методом усиления высказываний из S^{Ω} . Это ставит следующую задачу.

Задача 2. Разработать метод обнаружения закономерностей (1) на данных, представленных многосортными эмпирическими системами.

В настоящее время только методы индуктивного логического программирования (Inductive Logic Programming) могут работать с данными представленными в языке первого порядка эмпирическими системами, но эти методы не могут работать с сильно зашумлёнными данными. Поэтому потребовалось разработать оригинальный метод, работающий с данными представленными эмпирической системой \mathfrak{X} и обнаруживающий на ней множество высказываний вида (1) в логике первого порядка с вероятностной мерой. Этот метод изложен в [Витяев Е.Е., 1976, 2006; Kovalerchuk B., Vityaev E., 2000] и основан на семантическом вероятностном выводе, обладающем целым рядом важных свойств [Витяев Е.Е., 1976].

2.2. Построение количественной эмпирической теории (КЭТ) осуществляется на основании результатов теории измерений, дающих числовые представления величин и законов. В теории измерений найдены системы аксиом для большинства физических величин и фундаментальных физических законов [Krantz D.H. et al, 1971]. Если в ЛЭТ содержится какая-либо система аксиом теории измерений, то она дает числовые представления величин и функциональных зависимостей.

Проблема в построении КЭТ состоит в том, что далеко не для всех систем аксиом, которые могут быть обнаружены в ЛЭТ, существуют соответствующие им результаты теории измерений. Это ставит следующую задачу.

Задача 3. Найти обобщение теории измерений, которое бы позволило строить числовые представления величин и законов практически для любой системы аксиом.

Такое обобщение получено в работе [Витяев Е.Е., 1985] на основе теории конструктивных моделей [Ершов Ю.Л., 1980]. Значения величин в этом случае представляются натуральными, рациональными или другими эффективно вычислимыми числами (например, кодами). Теория конструктивных моделей наиболее полно отражает смысл построения числовых представлений – закодировать эмпирическую систему числами или кодами так, чтобы можно было легко и удобно вычислять значения истинности отношений и операций на эмпирической системе. Используя теорию конструктивных моделей, можно получить конструктивные числовые представления величин и законов практически для любой системы аксиом из ЛЭТ. Множество конструктивных числовых представлений будем называть *конструктивной эмпирической теорией* (КонЭТ).

2.3. Построение конструктивной эмпирической теории (КонЭТ). В теории измерений нельзя получить числовые представления для частичных порядков, толерантностей, решеток и т. д. В [Витяев Е.Е., 1985] приведены основные понятия конструктивных числовых представлений. На примере одной из наиболее распространенных систем аксиом экстенсивных величин в [Витяев Е.Е., 2006] доказано, что существует конструктивное числовое представление рациональных делений шкал приборов для этих величин. Примерами конструктивных числовых представлений законов являются, например, психологические тесты и анкеты.

2.4. Цикл компьютерного познания. Таким образом, разработаны понятия и методы, которые позволяют осуществлять следующий цикл познания, обозначенный на рис. 1 двойными стрелками:

- определить онтологию предметной области;
- извлечь из числовых представлений величин множество отношений и операций Ω , определяющих смысл величин в соответствии с теорией измерений. Представить систему понятий онтологии одноместными предикатами в Ω ;
- перевести данные в (многосортную) эмпирическую систему \mathfrak{X} , используя множество отношений и операций Ω ;
- перевести априорные и экспертные знания в систему аксиом S^{Ω} ;
- обогатить систему аксиом S^{Ω} путем обнаружения ЛЭТ, включая $\text{Th}(\mathfrak{Z})$ и множество

- закономерностей;
- найти системы аксиом теории измерений в S^{Ω} , которым удовлетворяют величины/законы;
- найти числовые представления величин/законов в теории измерений и/или в теории конструктивных моделей;
- проинтерпретировать полученные числовые представления в онтологии.

3. Реляционный подход к извлечению знаний.

Для компьютерной реализации приведенной схемы познания потребовалось разработать специальный метод обнаружения закономерностей в языке первого порядка с вероятностными оценками. Нами разработана программная система Discovery обнаружения закономерностей в языке первого порядка в условиях шумов. Данный подход назван нами как реляционный подход (Relational Data Mining) к извлечению знаний [Витяев Е.Е., Москвитин А.А., 1993; Kovalerchuk B., Vityaev E., 2000, 2008; E. Vityaev, B. Kovalerchuk, 2004; Витяев 2006].

Нами доказано [E. Vityaev, B. Kovalerchuk, 2004; Витяев 2006], что, в принципе, система *Discovery* позволяет обнаруживать на данных следующие множества закономерностей:

- a) теорию $Th(\mathcal{S})$;
- b) все закономерности, имеющие максимальные значения условной вероятности;
- c) все максимально специфические закономерности, включающие максимум контекста применения правила [Витяев 2006].

Таким образом, разработанный нами реляционный подход к извлечению знаний позволяет:

- (1) снять ограничения с используемых типов данных за счет применения теории измерений, онтологий и (многосортовых) эмпирических систем;
- (2) извлечь всю информацию из данных;
- (3) использовать любые априорные знания, представленные в языке первого порядка;
- (4) обнаружить любой класс гипотез (Rule Type) в языке первого порядка;
- (5) обнаружить системой *Discovery* закономерности вида (a), (b), (c);
- (6) обнаружить базу знаний, включающей множества закономерностей вида (a), (b), (c), которая будет полна в двух смыслах:
 - a) в смысле полноты извлечения информации из данных;
 - b) в смысле полноты обнаруживаемых множеств правил (a), (b), (c).

4. Применения реляционного подхода в финансовом прогнозировании, медицине и биоинформатике. Применения реляционного подхода в соответствии с циклом познания 2.4. дало следующие результаты:

I. Определить для используемых типов данных отношения и операции и преобразовать данные в многосортовые эмпирические системы:

- 1) в финансовых приложениях используются следующие функции и отношения, определяемые для временного ряда:
 - a) первая разность –

$$\Delta_{ij}(a_t) = (SP500C(a_t^j) - SP500C(a_t^i)) / SP500C(a_t^i), i < j, i, j = 1, \dots, 5$$
 Эта функция представляет собой разность между SP500C для i-х и j-х дней, нормализованных относительно SP500C для i-го дня;
 - b) разность между двумя относительными разностями $\Delta_{ijk}(a_t) = \Delta_{jk}(a_t) - \Delta_{ij}(a_t)$,
 - c) функция $wd(a)$ отображает пять календарных дней в числа.
 $wd(a) = \{1, 2, 3, 4, 5\}$ означает, что a представляет собой пять последовательных дней недели с понедельника по пятницу;
 - d) Отношение роста / падения цены с определенного дня недели по другой определенный день недели;
- 2) в приложениях по разработке диагностической системы рака груди использовались различные признаки определенные экспертом. Они включали в себя количественные, ранговые, номинальные и Булевы признаки;
- 3) в приложениях в биоинформатике использовались следующие операции и отношения, определяемые для первичных сигналов:
 - a) положение олигонуклеотидов относительно начала транскрипции;
 - b) взаимное расположение олигонуклеотидов в модели;
 - c) ориентация олигонуклеотидов в двойной спирали ДНК;
 - d) кроме того, сами сигналы могут быть достаточно разнообразны;

II. Используя найденные отношения и операции, определить класс гипотез Rule Type в

языке первого порядка для решения рассматриваемой прикладной задачи:

- 1) в финансах использовались следующие классы гипотез в терминах определенных отношений и операций, где Δ – любое из отношений $\Delta_{ij}(a_t) = \Delta_j(a_t) - \Delta_i(a_t)$, $\Delta_{ijk}(a_t) = \Delta_{jk}(a_t) - \Delta_{ij}(a_t)$, $i, j = 1, \dots, 5$:
 - a) множество гипотез H1 –

$$wd(a) = wd(b) = \langle d_1, \dots, d_5 \rangle \& (\Delta_{jk}(a) \leq \Delta_{jk}(b))^{\epsilon_1} \Rightarrow (S\&P500(a) \leq S\&P500(b))^{\epsilon_0}$$
 - b) множество гипотез H2 –

$$wd(a) = wd(b) = \langle d_1, \dots, d_5 \rangle \& (\Delta(a) \leq \Delta(b))^{\epsilon_1} \& (\Delta(a) \leq \Delta(b))^{\epsilon_2} \Rightarrow (S\&P500(a) \leq S\&P500(b))^{\epsilon_0}$$
 - c) множество гипотез H3 –

$$wd(a) = wd(b) = \langle d_1, \dots, d_5 \rangle \& (\Delta(a) \leq \Delta(b))^{\epsilon_1} \& (\Delta(a) \leq \Delta(b))^{\epsilon_2} \& (\Delta(a) \leq \Delta(b))^{\epsilon_3} \Rightarrow (S\&P500(a) \leq S\&P500(b))^{\epsilon_0}$$
 - d) Множество гипотез H4 –

$$wd(a) = wd(b) = \langle d_1, \dots, d_5 \rangle \& (\Delta(a) \leq \Delta(b))^{\epsilon_1} \& \dots \& (\Delta(a) \leq \Delta(b))^{\epsilon_k} \Rightarrow (S\&P500(a) \leq S\&P500(b))^{\epsilon_0}$$
- 2) в приложениях по разработке диагностической системы рака груди обнаруживались гипотезы, содержащие разнообразные признаки определенные экспертом;
- 3) в приложениях в биоинформатике обнаруживались так называемые комплексные сигналы вида:
 - a) $(S_1, \dots, S_{i-1}, S_i) = (\text{Позиция}(S_1) < \dots < \text{Позиция}(S_{i-1}) < \text{Позиция}(S_i))$, $i = 1, 2, \dots$

III. В результате проведенных экспериментов получены следующие выводы относительно применимости реляционного подхода в различных предметных областях:

- 1) применение в финансах показало:
 - a) система *Discovery* в состоянии обнаруживать закономерности в таких сильно зашумленных данных как финансовые ряды;
 - b) прогнозировать такие сложные данные как курсы акций и индексы, используя необычные отношения и операции;
 - c) получаемые правила интерпретируемы в финансовых терминах, что очень важно для таких ответственных областей, как финансы. Финансист с большим доверием будет вкладывать деньги, если он будет понимать используемые правила;
 - d) Многие люди за рубежом держат деньги в акциях и многие играют на них, используя самые разнообразные правила и индексы. Проверить же свои правила автоматически они не могут, так как нет методов, которые бы позволяли бы записывать и проверять разнообразные гипотезы. Опыт применения системы *Discovery* в финансах показал, что эта система может, в принципе, решить эту задачу;
- 2) применение в медицине показало, что можно извлечь из данных и эксперта совместное множество знаний для медицинской диагностической системы рака груди. Согласованная база знаний лишена противоречий между правилами, полученными системой *Discovery*, правилами, используемыми опытным радиологом, и базой данных патологически подтвержденных случаев;
- 3) Применение реляционного подхода в биоинформатике показало, что система *Discovery* может быть успешно использована для решения одной из сложнейших задач биоинформатики – анализа регуляторных районов генов. В отличие от других методов, система *Discovery* может быть применена иерархически к анализу различных уровней анализа генов.

ЛИТЕРАТУРА

- Витяев Е.Е.** Метод обнаружения закономерностей и метод предсказания. В кн.: Эмпирическое предсказание и распознавание образов (Вычислительные системы, вып. 67). Новосибирск, 1976, с. 54-68.
- Витяев Е.Е.** Конструктивное числовое представление величин. В кн. Методы анализа данных (Вычислительные системы, Вып. 111). Новосибирск, 1985, с.23-32.
- http://www.math.nsc.ru/AP/ScientificDiscovery/PDF/konstr_repres.pdf

- Витяев Е.Е.** Обнаружение закономерностей (методология, метод, программная система SINTEZ). 1. Методология // Методологические проблемы науки (Вычислительные системы, 138), Новосибирск, 1991, с. 26-60
http://www.math.nsc.ru/AP/ScientificDiscovery/PDF/pred_invention.pdf
- Витяев Е.Е.** Извлечение знаний из данных. Компьютерное познание. Модели когнитивных процессов: Моногр. // Новосибирский гос. ун-т. Новосибирск, 2006. 293с.
http://www.math.nsc.ru/AP/ScientificDiscovery/PDF/scientific_discovery.pdf
- Витяев Е.Е., Москвитин А.А.** Введение в теорию открытий. Программная система DISCOVERY. // Логические методы в информатике (Вычислительные системы, вып. 148), Новосибирск, 1993, с.117-163.
http://www.math.nsc.ru/AP/ScientificDiscovery/PDF/introducton_discovery_theory.pdf
- Ершов Ю.Л.** Проблемы разрешимости и конструктивные модели. - М.: Наука, 1980.-415с.
- Кини Р.Л., Райфа Х.** Принятие решений при многих критериях: предпочтения и замещения. М.:Радио и связь,1981.-560с.
- Кулаков Ю.И.** Новая формулировка теории физических структур // Методологические и технологические проблемы информационно-логических систем (Вычислительные системы, 125), Новосибирск, 1988, с.3-32.
- Лихорадка** // Малая медицинская энциклопедия, М.
- Михайличенко Г.Г.** Решение функциональных уравнений в теории физических структур // ДАН, т.206, N5, 1972, с.1056-1058.
- Пфанцгль И.** Теория измерений. Мир,1976. 248с.
- Clocksin W.F., Mellish C.S.** Programming in Prolog. NY, 1981, 280с.
- Kovalerchuk B., Vityaev E.** Data Mining in Finance: Advances in Relational and Hybrid Methods. Kluwer Acad. Publ, 2000, p.308
- Kovalerchuk, B., Vityaev E., Ruiz J.F.,** Consistent and Complete Data and "Expert" Mining in Medicine, In: Medical Data Mining and Knowledge Discovery, Springer, 2001, pp. 238-280.
http://www.math.nsc.ru/AP/ScientificDiscovery/PDF/ieee_medical_paper.pdf
- E. Vityaev, B. Kovalerchuk,** Empirical Theories Discovery based on the Measurement Theory. Mind and Machine, v.14, #4, 551-573, 2004
http://www.math.nsc.ru/AP/ScientificDiscovery/PDF/empirical_theories_discovery_based_on_the_measurement_theory.pdf
- B. Kovalerchuk, E. Vityaev,** Symbolic Methodology for Numeric Data Mining. *Intelligent Data Analysis*. Special issue on "Philosophies and Methodologies for Knowledge Discovery and Intelligent Data Analysis" eds. Keith Rennolls, Evgenii Vityaev. v.12(2), IOS Press, 2008, pp. 165-188
http://www.math.nsc.ru/AP/ScientificDiscovery/PDF/relational_methodology_paper.pdf
- Krantz D.H.,Zuce R.D., Suppes P., Tversky A.** Foundations of measurement.Vol.1. NY and London, Acad. press,1971. 577p.
- Website Scientific Discovery,** <http://www.math.nsc.ru/AP/ScientificDiscovery>