А Н Н О Т А Ц И Я

УДК 53.02+519.7+519.812.2

Существующие подходы к Теории Открытий в Искусственном Интеллекте связаны с разработкой вычислительных моделей исторического процесса открытий. Нами предлагается другой подход: не моделировать историю открытий, а выяснить как следует с современной точки зрения обнаруживать законы. Понятие Закона еще исследуется и эти исследования не закончены. Наш подход опирается на Теорию Измерений и Теорию Физических Структур. С точки зрения этих теорий существующие подходы к Теории Открытий являются аппроксимационными - вид закона должен быть определен априори. Существующие подходы к обнаружению зависимостей в Анализе Данных также являются аппроксимационными. Для обнаружения не аппроксимационных, а в определенном смысле "истинных" законов и адекватных шкал величин необходимо, согласно Теории Измерений и Теории Физических Структур разработать принципиально новый метод обнаружения "полного" множества закономерностей (включая системы аксиом Теории Измерений) в языке первого порядка на выборках из эмпирических систем. В работе анализируются принципы построения такого метода (сам метод описан ранее), стратегии исследования им данных и описание программной системы Discovery, реализующей этот метод.

ВВЕДЕНИЕ В ТЕОРИЮ ОТКРЫТИЙ. ПРОГРАММНАЯ СИСТЕМА DISCOVERY.

ВИТЯЕВ Е.Е., МОСКВИТИН А.А.

УДК 53.02+519.7+519.812.2

Введение: 1. В настоящее время в Анализе Данных и Искусственном Интеллекте разработано довольно много различных методов обработки данных. Однако, методология этих методов практически не разработана. Давно назрела потребность проанализировать эти методы с точки зрения их связи с процессом познания. Такому анализу и посвящено введение. Более подробно отдельные вопросы изложены в работах [1-8]. В результате анализа мы естественным образом придем к некоторому подходу к Теории Открытий, определенному ниже, разработанному в этих же работах, и программной системе Discovery, реализующей эту вполне определенную часть процесса познания. В процессе анализа основное внимание будет уделяться методам анализа зависимостей. Из методов Анализа Данных к методам анализа зависимостей относятся методы регрессионного и дискриминантного анализов, методы распознавания образов и обнаружения закономерностей, методы классификации и аппроксимации. Во всех этих методах вид зависимости задается внешним, априорным по отношению к самой зависимости образом. 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 ставится да и не может быть поставлен в рамках Анализа Данных. В Анализе Данных неизвестная зависимость аппроксимируется некоторыми заданными априори классами функций, моделями, решающими правилами и т.д. Аппроксимируя неизвестную зависимость с требуемой степенью точности и надежности, методы Анализа Данных решают по-существу задачу предсказания. Найденная аппроксимация ничего (или почти ничего см. #2 в [1]) не говорит о том какова "истинная" зависимость.

Процесс аппроксимации неизвестных зависимостей начинается с переноса способов измерения из точных наук и прежде всего из физики в другие области. Рассмотрим, например, такую физическую величину как температура (см. также]2 в [1]). 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 не только риски делений шкалы на приборе, а это прежде всего тот набор операций и отношений, которые имеет смысл производить с числовыми значениями величин с точки зрения рассматриваемой предметной области, точнее это те операции и отношения, которые интерпретируемы в системе понятий соответствующей предметной области. Некоторые возражают, что термометр не может измерять ничего кроме температуры. Он действительно во всех случаях измеряет физическую температуру. Но резонно в таком случае спросить, а зачем собственно мы измеряем температуру? Ведь не затем чтобы согласно законам физики узнать сколько в больном содержится тепла и сколько он в состоянии растопить льда, если его положить на лед, 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 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"[12]. Для почв температура должна интерпретироваться в системе понятий физиологии растений и деятельности микроорганизмов и т.д. Следует понимать, что физическая величина температуры является косвенным измерением некоторой другой величины, интерпретируемой в системе понятий предметной области, которую мы именно и хотим измерить. Физическая температура больного, например, есть косвенное измерение для 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 слепо переносить из одной области знаний в другую. После такого переноса необходимо заново определить шкалу. Например, для температуры больного интерпретируемы выделенные значения 36.7, 42. и отношение линейного порядка $<$, поэтому это будет шкала порядка с выделенными значениями.

На следующем шаге, при применении методов Анализа Данных к различным данным, также проявляется их аппроксимационный характер. Перед обработкой данные как правило преобразуются к одному из известных видов - количественным или качественным. Если они преобразуются к количественным данным (то есть с числами разрешается производить любые математические операции, вне зависимости от их интерпретации), то в них вносится бессмысленная информация (связанная с произволом в выборе числового представления), и как следствие проявляющаяся в том, что невозможно обоснованно проинтерпретировать полученные результаты. Если данные преобразуются в количественные за счет использования различного рода (числовых) моделей или дополнительных предположений, которые в этом случае не полностью интерпретируемы, то это также приводит к невозможности обоснованно проинтерпретировать полученные результаты. Если данные преобразуются в дискретные, то это ведет к потере информации. Поэтому неизвестные зависимости не просто аппроксимируются задаваемыми из вне видами зависимостей, но и сами данные часто искажаются, чтобы их обработка этими методами была возможна.

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 [8-11]. Теория Измерений основана на известном 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. 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 можно осмысленно проводить чтобы не получать бессмысленных результатаов типа 1.5 дровосека, 1м. + 1кг., и т.д. Эмпирическая система - это множество (идеализированных) объектов с заданным на нем множеством интерпретируемых в системе понятий отношений и операций, удовлетворяющих некоторой системе аксиом. Такой семантический уровень рассмотрения с необходимостью возникает из того факта, что интерпретировать человек может только качественно. Поэтому интерпретируя количественные значения величин, модели, функции и т.д. он интерпретирует их качественно - в системе понятий предметной области - и в качестве промежуточной стадии такой интерпретации - на семантическом уровне в (многосортной) эмпирической системе. Семантический уровень возникает не только из-за требования интерпретируемости, но он и исторически является первичным и представляет собой целостное (модельное) представление той исходной деятельности над объектами, которая привела в свое время к возникновению теорий и чисел [8-11].

Если в Анализе Данных зависимости аппроксимируются, то в Теории Измерений определяются в определенном смысле "истинные" величины и зависимости:

Числовые представления величин, получаемые из систем аксиом Теории Измерений, дают "истинные" шкалы величин, интерпретируемые в системе понятий соответствующей предметной области. "Истинные" шкалы это те, которые интерпретируются в системе понятий предметной области и являются числовыми кодами значений величины соответствующей эмпирической системы.

Числовые представления законов, получаемые на основании какой-либо системы аксиом, являются "истинными" законами данной предметной области в том смысле, что они, во-первых, интерпретируемы в системе понятий данной предметной области и являются числовыми кодами взаимосвязи величин из эмпирической системы и, во-вторых, их числовые представления либо получаются одновременно (единой процедурой шкалирования) с числовыми представлениями величин, либо согласованы с числовыми представлениями величин. В [8-11] показано, что уравнения для физических законов просты только потому, что они получаются процедурой одновременного шкалирования всех, входящих в зависимость величин, так чтобы взаимосвязь этих величин выражалась заданной, определяемой системой аксиом, простой функциональной зависимостью.

Следующий вывод, который следует из Теории Измерений, состоит в том, что числовой способ представления данных, используемый в Анализе Данных не является адекватным для целей обнаружения законов и зависимостей, хотя он вполне может быть адекватен для целей предсказания, где вид зависимости для нас не важен. Можно хорошо предсказывать, используя только двоичную кодировку данных. **Цель анализа зависимостей совсем другая - познать предметную область**. Для достижения этой цели **интерпретируемость данных и результатов анализа данных в системе понятий предметной области является необходимым условием получения хоть какого-нибудь полезного результата, вносящего вклад в теорию предметной области**. Так как числа сами по себе смысла не имеют, то интерпретируемость данных и результатов счета означает согласно Теории Измерений их интерпретируемость на семантическом уровне в системе понятий предметной области без использования чисел. Поэтому для целей анализа зависимостей необходим тот способ представления данных, который принят в Теории

Измерений - в виде (многосортных) эмпирических систем. Системы аксиом, которым удовлетворяют эти эмпирические системы, представляют собой **логическую эмпирическую теорию предметной области**. Системы аксиом как логические высказывания очевидно интерпретируемы в системе понятий предметной области. Поэтому анализ зависимостей должен состоять в обнаружении зависимостей на данных представленных (многосортными) эмпирическими системами. Каков при этом возможен вид зависимостей? Богатство языка логики первой ступени а также тот факт, что этот язык используется в Теории Измерений показывают, что этого языка вполне достаточно для выражения зависимостей. Таким образом, задача анализа зависимостей сводится к задаче усиления (в логическом смысле) логической эмпирической теории за счет обнаружения зависимостей в логике первого порядка. Числовые представления величин и функциональные зависимости для законов должны получаться из обнаруженных систем аксиом в результате применения Теории Измерений. Полученные шкалы величин и законы, связывающие величины, дают **Количественную Теорию Предметной Области**. Для физики этот переход продемонстрирован в [13] (см. также [4,8]). В [13] показано как можно строить Количественную Теорию Предметной Области - систему величин, связанных между собой (фундаменталь- ными) законами.

Таким образом, задача анализа зависимостей разбивается на два этапа: сначала надо построить Логическую Эмпирическую Теорию, а затем, применяя Теорию Измерений, 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. Для физики, например, этот процесс протекал достаточно долго.

Из всего сказанного следует, что Теория Измерений и должна быть той теорией, которую надо использовать для разработки методов обнаружения законов и шкал. Однако, точность аксиоматического анализа законов, проведенная в Теории Измерений, оборачивается сложностью применяемого ею аппарата, недостаточным разнообразием систем аксиом, чтобы можно было браться искать любые зависимости, отсутствием общего метода обнаружения систем аксиом в данных и другими проблемами. Прежде чем получить Теорию Открытий, основанную на Теории Измерений, необходимо заполнить эти пробелы или по крайней мере указать пути по которым эти пробелы могут быть заполнены. Для получения такой Теории Открытий необходимо было:

- 1.1. Разработать достаточно общий метод обнаружения систем аксиом на различных данных [2];

- 1.2. Найти обобщение Теории Измерений, позволяющее получать числовые представления практически для любых систем аксиом, которые может обнаружить этот метод [9-10];

- 1.3. Найти путь к построению систематик возможных законов природы [8].

Этим вопросам и посвящена серия работ [1-10], которая приводит в результате к логико-статистическому (и в этом смысле объективному в отличии от подходов к Теории Открытий в Искуственном Интеллекте см. п.2) продходу к Теории Открытий основанному на Теории Измерений и программной системе обнаружения систем аксиом Discovery.

Что можно сказать о возможностях построения Количественных теорий такой Объективной Теорией Открытий в других областях знания. Возможность построения Количественных Теорий в психологии, психофизике, социологии, психолингвистике, принятии решений, этике и т.д. подтверждают результаты, полученные в функциональной теории измерений [13,14]. Хотя методы функциональной теории измерений менее обоснованы и требуют более сильных и менее очевидных предположений, чем методы Теории Измерений (сравнение этих двух подходов обсуждено в [10]), тем не менее,

благодаря своей простоте, они позволили получить довольно много интересных результатов в этих областях. Только результаты, полученные в психологии, позволили их авторам утверждать, что "мышление человека, кажется, подчиняется довольно общей когнитивной алгебре".

2. Сравним предлагаемый подход с существующими подходами в Теории Открытий и Искусственном Интеллекте (см. обзор [19]). Цель этих подходов формулируется следующим образом: "…можно рассматривать программные системы открытий Искусственного Интеллекта как вычислительные модели исторического процесса открытий. "Моделировать исторический процесс открытия законов не зная что такое Закон вполне соответствует духу Искуственного Интеллекта ("Ньюелл и Саймон обосновали полезность достаточных моделей поведения" [19]). Но нами предлагается другой подход: не моделировать историю открытий, а выяснить как следует с современной точки зрения обнаруживать законы. Понятие Закона еще исследуется, и в самой физике, и в Теории Измерений, и в Теории Физических Структур [20-22] и в других работах (см., например, [4,8,23]) и эти исследования еще не закончены. Наш подход опирается на теорию Теорию Измерений и Теорию Физических Структур. С точки зрения этих теорий существующие подходы к Теории Открытий также являются аппроксимационными - вид закона должен быть определен априори. Из Теории Измерений также следует и то, что научившись моделировать процесс открытия законов на исторических физических примерах, нельзя будет применить их в нефизических областях. Для самой же физики эти методы не нужны. Как утверждается в [19] они, возможно, будут полезны для истории открытий: “В той степени, в какой шаги системы будут теми же, что и были сделаны исторически, можно будет сделать вывод, что эта система представляет собой подходящую модель исторических открытий".

3. Метод обнаружения "полного" множества интерпретируемых зависимостей (систем аксиом) в данных. Это одна из первых задач 1.1, которую необходимо было решить. Первый вариант этого метода практически аппробирован и опубликован в [2]. Приведем основную нить рассуждений, аргументы и ссылки, приводящую к построению метода и программной системе Discovery.

Из Теории Измерений следует, что эмпирическая теория предметной области может быть представлена многосортной эмпирической системой $M = \langle U,\{A_{i\in I}\};V,W\rangle$ и системой аксиом $S^{V\cup W}$, истинной на M, где U - генеральная совокупность объектов, $\{A_{i\in I}\}$ - множества объектов различных типов I (например, множество значений некоторой величины, наборы величин, связанных некоторой зависимостью, множество действительных чисел и т.д.), $V = \langle P_1,\ldots,P_k,r_1,\ldots,r_l\rangle$, - множество эмпирических отношений и операций, W - множество идеализированных отношений и операций, соответствующих отношениям из V. Отношения и операции словаря V эмпирической системы M эмпирически интерпретируемы, т.е. представляют собой некоторые измерительные процедуры (в том числе анкетирование, тестирование, экспертные оценки и т.д.) интерпретируемые (вместе с результатами) в системе понятий предметной области. Эмпирическая система M дает нам интерпретируемое семантическое представление предметной области, а система аксиом S^V - аксиоматическое описание предметной области.

В теории измерений существует два уровня рассмотрения - эмпирический и теоретический. На эмпирическом уровне рассматриваются реальные множества объектов $A \subset U$ (выборки из U), данные, представляемые эмпирическими системами $D = \langle A,V\rangle$ и системы аксиом S^V реальных приборов из V, а на теоретическом уровне - идеализированные эмпирические системы Теории Измерений $M = \langle A,W\rangle$ и системы аксиом S^W для идеализированных приборов из W и теоретические результаты о существовании, единственности и адекватности числовых представлений. Оба этих уровня вместе с взаимосвязью между ними с помощью правил соответствия [18] а также одновременное представление эмпирической системы, числовой системы и сильного гомоморфизма

эмпирической системы в числовую систему могут быть представлены в многосортной эмпирической системе M и системе аксиом $S^{V\cup W}$ за счет совместного представления реальных, идеализированных и числовых сортов. **Многосортную эмпирическую систему M** вместе с системой аксиом $S^{V\cup W}$ будем называть **эмпирической аксиоматической теорией**. Заметим, что в эмпирической аксиоматической теории числа играют вспомогательную роль и над ними можно производить только те математические действия, которые при обратном отображении относительно сильного гомоморфизма, преобразуются в определимые в словаре V отношения и операции.

В #4 [1] приводится методика извлечения из наиболее распространенных типов данных - парных сравнений, множественных сравнений, матричного представления бинарных отношений, матриц упорядочений и близости, матриц объект-признак – всей интерпретируемой в них информации и представления ее в эмпирических аксиоматических теориях. Извлеченная информация представляется в виде эмпирических систем и систем аксиом взятых как на эмпирическом так и на теоретическом уровнях. Как было сказано, без ограничения общности можно считать, что анализ зависимостей на данных, представленных эмпирическими аксиоматическими теориями, сводится к обнаружению зависимостей на эмпирической системе $M = \langle U,V \rangle$ в виде формул в языке первого порядка в словаре V (ввиду богатства этого языка). Такой анализ по-существу представляет собой аксиоматический анализ предметной области.

Зависимости в виде утверждений в языке первого порядка могут быть детерминированными (истинными на M) и не детерминированными. Детерминированные зависимости являются системами аксиом. Таким образом, "полный" анализ детерминированных зависимостей на M сводится к определению полного множества $S^V{}_M$ истинных на M формул в языке первого порядка.

Рассмотрим множество $S^V{}_M$. В #5 [1] аргументируется, что если ограничиться рассмотрением зависимостей, сформулированных только в виде универсальных формул (формул содержащие только кванторы всеобщности), то этого практически достаточно для формулировки зависимостей. Там же находится эмпирически интерпретируемое свойство измерительных процедур из которого следует универсальная аксиоматизируемость экспериментальной зависимости $S^V{}_M$. Таким образом, если это свойство выполнено для измерительных процедур из V (а оно является достаточно слабым и, как показано в #5 [1], практически не ограничивает "полноту" анализа зависимостей), то задача "полного" анализа детерминированных зависимостей сводится к определению множества $US^V(M)$ всех универсальных формул языка первого порядка, истинных на M. Известно, что любая конечная совокупность универсальных формул логически эквивалентна совокупности формул вида:

$$\forall x_1,\ldots,x_m(A_1\&\ldots\&A_n \Rightarrow A_0), \tag{1}$$

где $A_0,A_1,\ldots,A_n$ - атомарные формулы вида $P^\varepsilon(t_1,\ldots,t_k)$, $(t=g)^\varepsilon$; $t,g,t_1,\ldots,t_k$ - термы; $x_1,\ldots,x_m$ - переменные; $\varepsilon = 1(0)$ - определяет отсутствие (наличие) отрицания. Этот результат сводит задачу "полного" анализа детерминированных зависимостей на M к задаче нахождения множества $R^V[ul](M)$ всех формул вида (1), истинных на M.

Заметим, что формулы (1) "выделяют" из универсальных формул всю их способность к предсказанию - возможность предсказывать по условию формулы (1) заключение A_0. Такое выделение необходимо, так как зависимости нас интересуют ровно в той мере в какой они интерпретируемы и поддаются экспериментальной проверке и значит способны предсказывать. Для осуществления предсказаний более удобной является форма записи формул (1) через индивидные константы. Вместо кванторов всеобщности и связанных ими переменных введем константы $z_1,z_2,\ldots$. Тогда формулы вида (1) преобразуются в формулы вида:

$$A_1\&\ldots\&A_n \Rightarrow A_0, \tag{2}$$

где $A_0,A_1,\ldots,A_n$ - атомарные формулы вида $P^{\varepsilon}(t_1,\ldots,t_k)$, $(t=g)^{\varepsilon}$; $t(z^t_1,\ldots,z^t_{lt})$, $g(z^g_1,\ldots,z^g_{lg})$, $t_1(z^1_1,\ldots,z^1_{l1})$, … ,$t_k(z^k_{k1},\ldots,z^k_{lk})$ - термы; $z^t_1,\ldots,z^t_{lt}$, $z^g_1,\ldots,z^g_{lg}$, $z^1_1,\ldots,z^1_{l1}$, … , z^k_{k1}, … ,z^k_{lk} - индивидные константы; $\varepsilon = 1(0)$ - определяет отсутствие (наличие) отрицания. Смысл формулы (2) состоит в том, чтобы при любой фиксированной замене индивидных констант на объекты из некоторой модели (M или D) из истинности посылки следовала бы истинность заключения. Множество формул R^V[ul](M) после такого преобразования переходит в некоторое множество формул вида (2), истинных на M, которое обозначим через RC^V(M).

Для обнаружения недетерминированных зависимостей в языке первого порядка необходимо ввести вероятностную меру на формулах языка первого порядка. Такая мера может быть введена различными способами [19].

Реально мы никогда не имеем эмпирической системы M всей генеральной совокупности, а только лишь некоторую выборку (данные) D = <A,V>, $A \cup U$ из U представляющие собой подмодель многосорной эмпирической системы M. Искать зависимости можно только по данным D. Задача "полного" анализа зависимостей с учетом данного ограничения преобразуется в задачу "полного" анализа зависимостей на данных D. Взаимосвязь между этими двумя задачами устанавливается следующими двумя требованиями:

3.1. Нас интересуют такие зависимости на D (как детерминированные так и недетерминированные), которые обладают предсказательной способностью по отношению к другим случайно выбранным из U объектам.

3.2. "Полное" множество зависимостей на D (детерминированных и недетерминированных) должно в пределе (при увеличении выборки) стремиться к "полному" множеству зависимостей на M.

Нетрудно видеть, что $RC^V(M) \subset RC^V(D)$, так как $D \subset M$ - выборка D подмодель M. Поэтому условиям 3.1,3.2 для детерминированных зависимостей можно удовлетворить, если из множества RC^V(D) взять только те зависимости, которые в соответствии с определенным статистическим критерием, при некотором доверительном уровне α, давали бы требуемую надежность предсказания (оценку условной вероятности формулы (2)). Обозначим это множество через $RCP^V_{\propto}(D)$.

Рассмотрим недетерминированные зависимости и множество $\Im^A = \{A_{j\in J}\}$ всех атомарных высказываний в словаре $V \cup C$, где $C = \{c_{k\in K}\}$ - множество всех индивидных констант. Задачу "полного" анализа недетерминированных зависимостей на M определим как задачу нахождения всех наиболее сильных (дающих максимальную оценку условной вероятности) условных зависимостей между атомарными высказываниями из $\Im^A$. Нетрудно видеть, что такие зависимости должны удовлетворять следующим трем условиям:

3.3. $m(A_1,\ldots,A_n) > 0$ (в противном случае условная вероятность не определена);

3.4. Если из посылки $A_1,\ldots,A_n$ удалить одно или несколько атомарных высказываний, то условная вероятность $m(A_0/A_1,\ldots,A_n)$ уменьшится (то есть все атомарные высказывания существенны (повышают условную вероятность) для предсказания A_0);

3.5. Посылка полна - при добавлении к ней любого атомарного высказывания нарушается одно из первых двух условий, то есть либо посылка становится переопределенной ($m(A_1,\ldots,A_n) = 0$), либо ни одно из атомарных высказываний не может более увеличить оценку условной вероятности.

Зависимости, удовлетворяющие условиям 3.3-3.5, называются в [20-23] вероятностными закономерностями. Множество вероятностных закономерностей обозначим через REG[ularity](M). Определение вероятностной закономерности относится ко всей генеральной совокупности M (в предположении, что нам известна вероятность m). В [20-23] устанавливается связь между множеством детерминированных закономерностей RC^V(M) и множеством вероятностных закономерностей REG[ularity](M). Доказывается, что

детерминированные закономерности, удовлетворяющие некоторым естественным дополнительным условиям, которые логически не сужают множества $RC^V(M)$ (посылка не всегда ложна; при удалении какого-либо атомарного высказывания из посылки закономерность становится ложной; при замене заключения на ложь закономерность также становится ложной), являются частным случаем вероятностных закономерностей. Таким образом множество REG(M) является в этом смысле расширением множества $RC^V(M)$.

Для обнаружения вероятностных закономерностей по выборке D условия 3.3-3.5 проверяются для формул (2) с помощью определенных статистических критериев. Формулы вида (2), для которых эти условия критериев. Формулы вида (2), для которых эти условия с некоторым уровнем доверия α выполнены, назовем закономерностями. Множество закономерностей, имеющихся на D, обозначим через $RG^V_\alpha(D)$. В силу того, что детерминированные закономерности являются в определенном смысле частным случаем вероятностных закономерностей, то обнаружение множества детерминированных закономерностей $RCP^V_\alpha(D)$ происходит автоматически при обнаружении $RG^V_\alpha(D)$. Во множестве закономерностей $RG^V_\alpha(D)$ детерминированные закономерности выделяются тем, что они не имеют исключений (при истинности посылки заключение всегда истинно). Таким образом, задача "полного" анализа зависимостей сведена к задаче обнаружения множества закономерностей $RG^V_\alpha(D)$. Эта задача решается методом обнаружения закономерностей, который приводится в [2].

Метод обнаружения закономерностей реализован программной системой Discovery.

Пользователь в диалоге с программной системой задает некоторое параметрическое семейство формул:

$$A_1\&\ldots\&A_n \Rightarrow A_0 \qquad (3)$$

где $A_0,A_1,\ldots,A_n$ - логические выражения (логические выражения включают логические связки AND, OR, NOT, скобки и произвольные арифметические выражения с параметрами). Параметры могут быть произвольными и изменяться в цикле. Параметрами могут быть номера признаков, интервалы изменения признаков, выделенные значения признаков, параметры модифицирующие признак (подвергающие его различным преобразованиям) и т.д. Для каждого набора значений параметров мы получаем конкретную формулу вида (3), которая проверяется на закономерность на многосортной эмпирической системе D.

#1. Стратегии поиска "полного" множества закономерностей.

1. Цель каждой стратегии - получить по возможности "полное" множество зависимостей в данных. Полный перебор всех правил вида (3), как правило, невозможен (только для отдельных типов данных, например, булевых или наименований он в принципе возможен, и то, если ограничить его требуемой надежностью предсказания). Поэтому необходим направленный перебор. Направленный перебор должен преследовать следующие цели:

1.1.1. Не пропустить статистически значимые закономерности;

1.1.2. Среди статистически значимых закономерностей найти наиболее точные в смысле максимальности полноты условий в посылке и обеспечивающих максимальную оценку предсказания (см. Ниже определение отношения $\succ$ "быть более точным правилом");

1.1.3. Среди закономерностей, удовлетворяющих первым двум условиям, т.е. статистически значимых и наиболее точных, найти закономерности, использующие наиболее тонкие свойства шкалы и соответсвующие им отношения и операции, которые с максимальной точностью уже содержательной, учитывающей интерпретацию и содержательную взаимосвязь отношений и операций, выражают зависимость в данных. Как правило такие закономерности более точны и в смысле п.1.2 (имеют большую оценку условной вероятности), хотя и несравнимы по отношению $\succ$.

Программная система DISCOVERY позволяет реализовывать направленный перебор с помощью стратегий направленного и все более точного анализа эмпирического содержания данных. При этом точность в соответствии с п.1.2 и п.1.3 понимается в двух смыслах, в смысле отношения $\succ$ и в смысле богатства информации. Шкалы величин упорядочены в соответствии с богатством информации, содержащейся в значениях величин - от шкал наименований и шкал порядка к шкалам интервалов, отношений и абсолютным шкалам. В программной системе уточнения правил осуществляются следующим образом:

1.1.4. Путем добавления новых условий в посылку, либо применением подстановок. Это может делаться самим исследователем или экспертом, опираясь на свою интуицию и определение отношения $\succ$ - "быть более точным правилом", введенного и исследованного в работах [18,19];

1.1.5. Такое уточнение кроме того, как показано в работах [18,19], будет представлять собой некоторый вероятностный вывод, определяющий поиск наиболее точного (имеющего максимальную оценку условной вероятности) правила.

1.1.6. Путем использования более "тонких" свойств величин, представленных соответствующими операциями и отношениями более сильных шкал;

Уточнения 1.1.4 - 1.1.6 позволяют удовлетворить соответствующим целям 1.1.2, 1.1.3. Цель 1.1.1 - не пропустить статистически значимой закономерности - в практически достаточной мере достигается тем, что для реальных задач, как правило, всегда выполняется условие: если есть закономерность со сложной посылкой, то такую закономерность всегда можно получить в результате последовательного уточнения посылки, начиная с простейшей, так, что все последующие уточнения так же будут закономерностями (эта процедура на генеральной совокупности представляет собой вероятностный вывод. Этот вероятностный вывод введен и исследован в работах [28-29]). Кроме того, для достижения этой же цели следует начинать с обнаружения "полного" множества закономерностей сначала на бедных шкалах булевых и наименований, а затем уже переходить к более сильным шкалам, т.е. стратегия должна опираться на упорядоченность шкал по богатству информации (логической силе). Это следует из того, что для бедных шкал в большей мере возможен полный перебор с которого и начинается процесс последовательного уточнения закономерностей.

Таким образом, для того чтобы получить стратегию нужно следовать пунктам 1.1.4 - 1.1.6. Но кроме того следует учитывать упорядоченность шкал, результаты Теории Измерений и те типы данных, которые у Вас имеются. В соответствии с этим ниже конкретно описаны различные возможные стратегии.

Приведем определение отношения $\succ$ - "быть более точным" [28]. Обозначим множество всех подстановок не являющихся перестановками через Qt (тождественная подстановка принадлежит Qt).

ОПРЕДЕЛЕНИЕ. Отношение $C \succ C'$, $C = A_1\&\ldots\&A_n \Rightarrow A_0$; $C' = A'_1\&\ldots\&A'_{n'} \Rightarrow A'_0$, $n,n' > 0$ имеет место тогда и только тогда, когда существует подстановка $\theta \in \Theta$ такая, что $A_0\theta = A'_0$ и $\{A_1\theta\&\ldots\&A_n\theta\} < \{A'_1,\ldots,A'_{n'}\}$ и либо θ не тождественная подстановка, либо $n < n'$.

Прежде чем привести стратегии, необходимо описать типы данных (не шкалы), встречающиеся в различных предметных областях. Эти типы данных определяют множество отношений и операций, которые для этих данных определены, но не аксиомы, которым эти отношения и операции удовлетворяют. Типы данных могут быть следующими: логический, наименований, порядка и отношений. Такими типами данных могут быть, соответственно, булевы матрицы $L(x_1,\ldots,x_n)$, со значениями во множестве $\{0,1\}$; матрицы в шкале наименовоний $N(x_1,\ldots,x_n)$, со значениями во множестве натуральных чисел; матрицы в шкале порядка $P(x_1,\ldots,x_n)$, с порядковыми значениями (числовые значения сравниваются только отношением порядка) и матрицы в шкале отношений (интервалов, абсолютной)

$A(x_1,\ldots,x_n)$, со значениями во множестве вещественных чисел. Здесь $x_1,\ldots,x_n$ - переменные не только по объектам, но и по признакам, номерам экспериментов, числу градаций признаков и т.д. Мы всегда будем считать, что у нас есть переменные двух типов - переменные по объектам из множества $\{a_1,a_2,\ldots\}$ и переменные-параметры из множества $\{par_1,par_2,\ldots\}$. Каждый параметр есть переменная, пробегающая некоторое конечное множество натуральных чисел. Эти множества могут быть различны для различных переменных. Всегда далее будем иметь ввиду, что на соответствующих местах в матрице могут стоять переменные только строго определенного типа. Кортеж переменных будем обозначать через $<x>$. Аксиомы, которым должны удовлетворять эти отношения и операции, должны быть получены уже в результате анализа этих данных, например, системой DISCOVERY.

1.1. В соответствии с упорядоченностью шкал, сначала следует провести обработку данных в шкале наименований. Имеющиеся числовые значения следует разбить на интервалы, которые в формуле (3) можно задать параметрами. Приведем общий вид закономерностей такого вида. Чтобы формула была легко обозримой мы будем всячески избегать индексов.

$$\begin{aligned}
&\forall a(L^{\varepsilon}(<x>)\&\ldots\&L^{\varepsilon}(<x>)\& \\
&\quad (N(<x>) = FN)^{\varepsilon}\&\ldots\&(N(<x>) = FN)^{\varepsilon}\& \\
&\quad (F < P(<x>) < F)^{\varepsilon}\&\ldots\&(F < P(<x>) < F)^{\varepsilon}\& \\
&\quad (F < A(<x>) < F)^{\varepsilon}\&\ldots\&(F < A(<x>) < F)^{\varepsilon} \Rightarrow \qquad (4) \\
&\quad \{L\,(<x>) \mid (N(<x>) = FN)^{\varepsilon} \mid (F < P(<x>) < F)^{\varepsilon} \mid (F < A(<x>) < F)^{\varepsilon}\},
\end{aligned}$$

где

a - одна переменная по объектам, входящая в кортежи $<x>$;

$w \in \{0,1\}$ - определяет отсутствие или наличие отрицания;

FN - целочисленные арифметические выражения, не содержащие переменной по объектам и включающие произвольное число переменных-параметров;

F - арифметические выражения, не содержащие переменных по объектам и включающие переменные-параметры;

$\mid$ - означает, что в заключении правила (4) стоит только одно из указанных в фигурной скобке выражений.

Закономерности вида (4), обнаруженные для количественных переменных, дают качественный ответ на вопрос, а существует ли вообще количественная зависимость. Если для некоторой зависимости (4) в дальнейшем будет найдена более точная зависимость и в смысле прогноза (оценки условной вероятности) и в смысле вида (логически более сильная) зависимость, то первую зависимость можно будет удалить.

1.2. Для отношения порядка и для сочетания типов наименований и порядка существует несколько различных классов закономерностей.

1.2.1. Прежде всего это аксиоматический анализ самого отношения порядка - определить какой системе аксиом отношение порядка удовлетворяет. Само отношение порядка часто задается матрицами бинарного отношения типа L(a,b). В настоящее время известно несколько десятков различных систем аксиом для отношений порядка. Программная система DISCOVERY содержит Help систем аксиом в котором есть достаточно большое число систем аксиом для отношений порядка. Они довольно разнообразны. Их следует просто проверить.

1.2.2. Закономерности между отношениями порядка – участки монотонной зависимости. Монотонные зависимости как правило не распространяются на все множество объектов, поэтому необходимо находить участки монотонной зависимости, ограничивая области монотонности формульно определяемыми одноместными предикатами. Приведем общий вид таких закономерностей. Определим сначала формульно определимые одноместные предикаты. Обозначим посылку правила (4) через F(a).

$$F(a) \equiv (L^{\varepsilon}(<x>)\&\ldots\&L^{\varepsilon}(<x>)\&$$

$(N(<x>) = FN)^{\varepsilon}\&\ldots\&(N(<x>) = FN)^{\varepsilon}\&$
$(F < P(<x>) < F)^{\varepsilon}\&\ldots\&(F < P(<x>) < F)^{\varepsilon}\&$ (5)
$(F < A(<x>) < F)^{\varepsilon}\&\ldots\&(F < A(<x>) < F)^{\varepsilon}$

Формула F(a) определяет некоторый участок (область) в обобщенном признаковом пространстве, включающем признаки разных шкал, относительно переменной по объектам a. Сформулируем сначала свойство монотонности непосредственно для отношений порядка ≤, >, которые могут быть заданы, например, типом данных L(a,b,n) или P(a,b,n), где n - номер признака, эксперимента или эксперта.

$$\forall a,b(F(a)\&F(b)\&(a \leq_{i1} b)^{\varepsilon}\&\ldots\&(a \leq_{im} b)^{\varepsilon} \Rightarrow (a \leq_{i0} b)^{\varepsilon}), \quad (6)$$

где i1,…,im,i0 - номера признаков;
$\varepsilon \in \{0,1\}$, $(a < b)^0 = (a > b)$, $(a < b)^1 = (a < b)$.

Напомним, что мы всячески избегаем индексов и поэтому w может принимать различные значения для различных отношений и вместо отношения ≤ в формуле (6) в произвольных местах может стоять отношение <. Отношение порядка также часто задается баллами посредством типа данных N(a,b,n).

1.2.3. Системы аксиом для законов в виде простых полиномов [13]. В [13] не только получены системы аксиом для основных физических законов, но и разработан целый класс систем аксиом для законов в виде простых полиномов. Особенность этого класса в том, что системы аксиом сформулированы относительно простейших отношений - отношений равенства или эквивалентности для всех величин и только для одной из величин - отношение линейного порядка. Если эти отношения удовлетворяют одной из разработанных там систем аксиом, то отсюда следует, что существуют сильные (лог-интервальные) шкалы для всех величин, связанные между собой заданным полиномом. Шкалы и полином получаются процедурой одновременного шкалирования. Эти результаты показывают, что для получения сильных шкал и законов не нужны какие-то особые отношения и операции для величин, а нужна прежде всего определенная их взаимосвязь. Эти результаты подтверждают возможность построения Количественных Теорий в различных предметных областях. Как было отмечено во введении, при переносе измерительных процедур из одной области знания в другую множество интерпретируемых отношений и операций следует пересматривать. Как правило их становится меньше, шкалы величин обедняются и бытует мнение, что тем самым какая-то важная информация теряется. Упомянутые выше результаты показывают, что это заблуждение, т.к. отношения равенства, эквивалентности и порядка при этом, как правило, всегда остаются. Используя их и определив систему аксиом, связывающую эти отношения, можно, в принципе, построить новые более адекватные для данной предметной области сильные шкалы величин и определить связывающий их закон. Поэтому дело не в богатстве отношений, а в их закономерных связях. Мы не будем приводить здесь эти системы аксиом, просто сформулируем в наиболее общем виде получаемые из них шкалы и законы.

Для любого простого полинома y = f(x1,…,xn) существует (либо может быть построена) система аксиом S в словаре $W = < \leq_y, =_{x1},\ldots,=_{xn}>$, из истинности которой на некоторой эмпирической системе, вытекает существование числовых представлений (сильных гомоморфизмов относительно не только отношений и операций словаря W, но и некоторых определимых через них отношений и операций) $\varphi_y:Y \to Re$; $\varphi_{x1}:X \to Re;\ldots;$ $\varphi_{xn}:X \to Re$ величин y,x1,…,xn, связанных данным простым полиномом $\forall r(\varphi_y(y) = f(\varphi_{x1}(x1),\ldots,\varphi_{xn}(xn)))$, r = <y,x1,…,xn>.

1.2.4. Суммируем стратегии для шкал порядка. Если есть отношения порядка, системы аксиом которых неизвестны, то их следует установить. Если какая-то система аксиом окажется выполненной, то там же будет указано какой это порядок и какое числовое или конструктивное числовое [9] представление для него существует, либо таковое неизвестно. Если числовое представление неизвестно, то можно попытаться самому построить конструктивное числовое представление как это определено в [9].

Если среди величин есть отношение линейного порядка, то следует проверить системы аксиом для простых полиномов. Если какая-то из систем аксиом для простых полиномов окажется выполненной (как при этом обходиться с аксиомами содержащими кванторы существования сказано в [2]), то это укажет способ совместного шкалирования всех входящих в зависимость величин так, что они будут связаны законом в виде соответствующего простого полинома. Это даст нам новые адекватные шкалы величин и связывающий их закон.

Если все порядки линейные, но ни одна из систем аксиом для простых полиномов не выполнена, то следует провести аксиоматический анализ свойств функциональной зависимости. Для шкал порядка это - свойство монотонности, приведенные в п. 1.2.2. Общий вид таких зависимостей содержится в HELP-е.

Если ни одна из упомянутых выше систем аксиом, содержащихся в HELP-а не выполнена, то можно продолжить аксиоматический неизвестной зависимости как это указано в п. 1.1.4. Для этого надо самому исследователю или эксперту, опираясь на свою интуицию, строить все более точные параметрические семейства правил, уточняющие закономерности в соответствии с определением отношения ➢.

1.3. Рассмотрим типы данных L(a,b,c,d) и P(a,b). Тип данных P(a,b) называется матрицей близости (см. [2] или обзоры [20,21]). В [2] показано, что этот тип данных может быть представлен типом данных L(a,b,c,d). Пусть дано некоторое множество объектов $A = \{a_1,\dots,a_m\}$. Матрицей близости для этих объектов называется матрица (r_{ij}), $i,j = 1,\dots,m$; r_{ij} - числовые оценки меры близости (сходства или различия) в порядковой шкале (имеют смысл только сравнения величин $r_{ij} \le r_{kl}$). Такие матрицы возникают, например, при сравнении или оценке экспертом всех пар объектов из A в некотором отношении.

Матрицы близости обрабатываются методами многомерного неметрического шкалирования [20,21]. Целью этих методов является представление объектов точками в некотором метрическом пространстве (Евклидовом или Римановом) минимальной размерности так, чтобы расстояния d_{ij} между ними с точностью до порядка соответствовали бы величинам r_{ij}. После применения методов многомерного шкалирования получается представление данных в метрическом пространстве после чего, как правило, применяются методы Анализа Данных. Недостатки такого подхода обсуждены в [2].

Чтобы представить эти данные в терминах Теории Измерений определим на множестве A отношение P типа L(a,b,c,d):

$P(a,b,c,d) \Leftrightarrow r_{ab} \le r_{cd}$.

В Теории Измерений эмпирические системы, включающие подобные четырехместные отношения, обозначаются как $M = \langle A^*;\le\rangle$, где $A^* \subset AxA$, $\le$ - бинарное отношение упорядочения, определенное на A^* и называются шкалами разностей [13,2] (положительных разностей, алгебраических разностей, равных конечных промежутков, абсолютных разностей и т.д.).

Следует особо подчеркнуть, что несмотря на то, что разница между шкалой порядка и шкалой разностей небольшая - в обоих случаях мы имеем только одно отношение порядка, одно из которых определено на объектах, а другое на парах объектов, - тем не менее в первом случае мы имеем только шкалу порядка, а во втором случае уже сильную шкалу - шкалу отношений (как и для простых полиномов). Например, для шкалы положительных разностей [13,2] существует шкала Ф такая, что для любых пар $(a,b),(b,c),(c,d) \in A^*$:

1.3.1. $(a,b) \le (c,d) \Leftrightarrow \Phi(a,b) \le \Phi(c,d)$.

1.3.2. $\Phi(a,c) = \Phi(a,b) + \Phi(b,c)$.

Отображение Φ единственно с точностью до положительного множителя (шкала отношений). Разница в том, что, имея возможность сравнивать отрезки, можно проводить различные процедуры шкалирования, откладывая равные отрезки. В Теории Принятия Решений известны десятки процедур одномерного шкалирования для субъективных оценок Лица Принимающего Решение. Эти же процедуры можно применять и для перешкалирования физических приборов, применяемых за пределами физики.

Для проверки систем аксиом разностей необходимо проверить, имеющиеся в HELP-е системы аксиом.

1.4. Как уже было сказано, отношения порядка, эквивалентности и равенства, как правило, всегда остаются при применении физических приборов за пределами физики и получают новую интерпретацию в системе понятий соответствующей предметной области. Поэтому системы аксиом для простых полиномов, шкал разностей и шкал порядка являются основными при пересмотре шкал и обнаружении законов в новых предметных областях.

1.5. Системы аксиом многих физических величин основаны на том факте (хотя есть довольно много исключений см. [13]), что для этих физических величин существует эмпирически интерпретируемая операция •, обладающая свойствами операции сложения. Для времени, например, эта операция определяет время $t_1 \bullet t_2$ равное двум следующим друг за другом промежуткам времени t_1 и t_2; для массы эта операция $a_1 \bullet a_2$ интерпретируется как совместное взвешивание двух объектов a_1 и a_2; для длин она интерпретируется как длина $l_1 \bullet l_2$ равная общей длине двух, положенных вдоль одной прямой, отрезков l_1 и l_2. Такие величины называются экстенсивными. Если для какой-либо из величин в данных есть интерпретируемая двуместная операция •, то, используя системы аксиом экстенсивных структур, можно проверить являются ли эти величины экстенсивными. Для экстенсивных величин можно построить шкалу отношений, такую же как и для упомянутых физических величин.

1.6. Все физические величины (как показано в [13])
разбиваются на две группы: те для которых существует операция • и те для которых такой операции нет. Те величины, для которых такой операции нет, получают сильную щкалу за счет законов, связывающих эту величину с величинами, имеющими сильные шкалы. Системы аксиом таких законов приведены в [13] (см. также [4]). Они описывают законы вида $y^n = x^m * z^k$ и кроме того дают сильные шкалы для величин, у которых нет операции •. Почти все простейшие физические законы, связывающие между собой всю совокупность физических величин в единую систему физических величин, имеют именно такой вид [13]. За счет этих законов все физические величины имеют сильные шкалы. Проверить эти системы аксиом можно системой Discovery.

1.7. Если ни одна из систем аксиом пункта п.1.6. не выполннена, а операции * и + для некоторых из величин интерпретируемы, то можно исследовать свойства функций путем последовательного уточнения правил с помощью отношения ➢, как указано в п.1.1.4. Поскольку функции исследованы достаточно хорошо, то можно предложить сразу несколько вариантов дальнейшего уточнения вида функций. Уточнения, в частности, могут делаться для выделенных в п. 1.2.2 участков монотонности.

1.7.1. Приведем сначала целый класс свойств, которые обычно не исследуются, а предполагаются заданными априори – свойства метрики для образов, кластеров или функций:

$\forall a,b\{F(a)\&(NO(a) = FN)\&(|x(b)\text{-}x(a)| \le A)\&\ldots\&$

$(|x(b)\text{-}x(a)| \le A) \Rightarrow (NO(b) = FN)\}$, где (7)

F(a) - формула (5);

x(a),x(b) - (разные) признаки объектов a,b;

| - | - некоторая метрика или норма;

NO - признак номера образа;

FN - целочисленное арифметическое выражение.

1.7.2. Приведем простейшее свойство для классов функций - линейность функции (интерпретировать его лучше как свойство пропорциональных изменений для некоторых областей):

$\forall$a,b{F(a)&F(b)&((x(b) - x(a)) $\leq A_1$) $\Rightarrow$ (8)

($A_3 \leq A_2$ *(y(b) - y(a))/(x(b) - x(a)) $\leq A_4$)}, где

F(a),F(b) - формулы (5);

A_1,A_2,A_3,A_4 - арифметические выражения с параметрами.

1.7.3. Участки монотонности с допустимыми отклонениями.

$\forall a,b(F(a)\&F(b)\&(a \pm A_1 \leq_{i1} b)^{\varepsilon}\&\ldots\&(a \pm A_k \leq_{ik} b)^{\varepsilon} \Rightarrow (a \pm A_0 \leq_{i0} b)^{\varepsilon}$,

где F(a) и F(b) - формулы вида (5);

i1,…,ik,i0 - номера признаков; (9)

$A_1,\ldots,A_k,A_0$ - арифметические выражения с параметрами;

$\varepsilon \in \{0,1\}$, $(a \pm A \leq_i b)^1 = (a \pm A \leq_i b)$, $(a \pm A \leq_i b)^0 = (a \pm A >_i b)$.

Нетрудно видеть, что продолжать уточнять вид функциональной зависимости можно самыми разнообразными способами и привести их все здесь просто невозможно. Список видов зависимостей может постоянно пополняться самими пользователями, поэтому в процессе эксплуатации он будет непрерывно обогащаться.

Укажем теории, на которые можно опираться при анализе свойств функций. Известно, что любую непрерывную функцию можно приблизить рядом Тейлора. В качестве свойств функции можно исследовать поведение приближений к производным. Есть функции задающиеся дифференциальными уравнениями. Для них можно исследовать соответствующие этим уравнениям приближения в виде разностных уравнений. Только на разностные уравнения нужно смотреть не как на способ приближенного вычисления, а как на свойства этих функций.

Может показаться, что исследование свойств функций ничем не отличается от тех аппроксимаций, которые находятся методами Анализа Данных. Напомним, что есть два принципиальных отличия: во-первых, свойства функций всегда интерпретируемы в системе понятий предметной области (только такие и рассматриваются); и, во-вторых, свойства функций могут быть установлены со сколь угодно большой степенью достоверности. Это позволяет говорить о них как об "истинных", а не как о аппроксимациях.

Л И Т Е Р А Т У Р А

1. Витяев Е.Е. Обнаружение закономерностей (методологоя, метод, программная система SINTEZ). 1. Методология // Методологические щроблемы науки (Вычислительные системы, 138), Новосибирск, 1991, с. 26-60

2. Витяев Е.Е. Метод обнаружения закономерностей и метод предсказания // Эмпирическое предсказание и распознавание образов (Вычислительные системы, вып. 67), Новосибирск, 1976, с. 54-68.

3. Витяев Е.Е. Закономерности в языке эмпирических систем // Эмпирическое предсказание и распознавание образов (Выч. сист., вып. 76), Новосибирск, 1978, с. 3-14.

4. Витяев Е.Е. Закономерности в языках эмпирических систем и законы классической физики // Эмпирическое предсказание и распознавание образов (Вычислительные системы, вып. 79), Новосибирск, 1979, с. 45-56.

5. Витяев Е.Е. Обнаружение функциональных зависимостей с одновременным формированием понятий // Вторая всесоюзная конференция по автоматизации поискового конструирования, Новосибирск, 1980, с. 171-172.
6. Витяев Е.Е. Упрощение функциональных зависимостей за счет перешкалирования величин // 11-я Всесоюзная школа-семинар по "Программно-алгоритмическому обеспечению прикладного многомерного статистического анализа", М., 1983, с.260-262.
7. Витяев Е.Е., Москвитин А.А. ЛАДА - программная система логического анализа данных // Методы анализа данных (Вычислительные системы вып.111), Новосибирск, 1985, с.38-58.
8. Витяев Е.Е. Числовое алгебраическое, и конструктивное представление одной физической структуры // Логиго-математические основы МОЗ (Вычислительные системы, вып. 107), Новосибирск, 1985, с.40-51.
9. Витяев Е.Е. Конструктивное числовое представление величин // Методы анализа данных (Вычислительные системы, вып. 111), Новосибирск, 1985, с.23-32.
10. Витяев Е.Е. Шкала экстенсивных величин как абстрактный тип данных // Всесоюзная конференция по прикладной логике (Тезисы докладов), Новосибирск, 1985, с.37-39.
11. Витяев Е.Е. Логико-операциональный подход к анализу данных // Комплексный подход к анализу данных в социологии. Тр. Института Социологических исследований АН, Москва, 1989, с. 113-122.
12. Лихорадка // Малая Медицинская Энциклопедия, М., 13. Психологические измерения / Под ред. Л.Д.Мешалкина. - М.: Мир, 1967. - 120с.
14. Foundations of measurement. Vol. 1. / Krantz D.H., Luce R.D., Suppes P., Tversky A. - NY and London: Academic press, 1971. - 577p.
15. Пфанцагль И. Теория измерений. М., Мир, 1976. - 248с.
16. Фишберн П.С. Теория полезности для принятия решений. - М., Наука, 1978. - 352с.
17. Anderson N.H. Integration theory, functional measurement and the psychologycal law // Advances in psychophysics./Ed. Geissler, Yu.Zabrodin, Berlin, 1976, p.93-130.
18. Anderson N.H. Algebraic Rules in Psyhological measurement. - Amer.Scientist, 1979 , v.67, p. 555-563.
19. Langley P., Zytkow J.M. Data-Driven Approaches to Empirical Discovery // Artificial Intelligence, v.40 (1989), N.1-3, p.283-312.
20. Кулаков Ю.И. Новая формулировка теории физических структур // Методологические и технологические проблемы информационно-логических систем (Вычислительные системы, 125), Новосибирск, 1988, с.3-32.
21. Самохвалов К.Ф. К обоснованию теории физических структур // Методологические и технологические проблемы информационно-логических систем (Вычислительные системы, 125), Новосибирск, 1988, с.33-41.
22. Михайличенко Г.Г. Решение функциональных уравнений в теории физических структур // ДАН, т.206, N5, 1972, с.1056-1058.
23. Витяев Е.Е., Костин В.С. Естественная классификация как закон природы // Интеллектуальные системы и методология. ("Материалы научно-практического симпозиума "Интеллектуальная поддержка деятельности в сложных предметных областях", Новосибирск, 7-9 апреля 1992), вып. 4, Новосибирск, 1992. с.107-115.
24. Pzelecki M. The logic of empirical theories. - London: Routledge Kogan Paul, 1969. - 109p.
25. Halpern J.Y. An Analysis of First-Order Logic of Probability // Artificial Intelligence v.46, 1990, p.311-350.
26. Витяев Е.Е. Обнаружение закономерностей, выраженных универсальными формулами // Эмпирическое предсказание и распознавание образов (Вычислительные системы, вып. 79), Новосибирск, 1979, с. 57-59.

27. Витяев Е.Е. Классификация как выделение групп объектов, удовлетворяющих разным множествам согласованных закономерностей // Анализ разнотипных данных (Вычислительные системы вып. 99), Новосибирск, 1983, с. 44-50.
28. Prediction and inductive synthesis of PROLOG-programs by a probabilistic model of data. Предсказание и индуктивный синтез ПРОЛОГ-программ по вероятностной модели данных // Institute of Mathematics SD AS USSR, Тр. ИМ СО АН СССР, 1990.
29. Витяев Е.Е. Семантический подход к созданию баз знаний. Семантический вероятностный вывод наилучших для предсказания ПРОЛОГ-программ по вероятностной модели данных. // Логика и семантическое программирование (Вычислительные системы, вып. 146), Новосибирск, 1992, с.19-49.
30. Каменский В.С. Модели и методы неметрического многомерного шкалирования (Обзор) // Автоматика и телемеханика, 1977, #8, с.118-156.
31. Терехина А.Ю. Методы многомерного шкалирования и визуализации данных (Обзор). - Автоматика и телемеханика, 1973, #7, с.80-94.
32. Косарев Ю.Г., Москвитин А.А. Проблемно-инструментальная технология построения программных систем // Методы анализа данных (Вычислительные системы, вып. 111), Новосибирск, 1985, с.59-75.