

## НЕЛИНЕЙНАЯ ДИНАМИКА МЫШЛЕНИЯ ЭКСПЕРТА И АЛГОРИТМЫ РАСШИРЕНИЯ ЕГО ПРОФЕССИОНАЛЬНЫХ ЗНАНИЙ НА ПРИМЕРЕ МЕДИЦИНСКИХ ЗАДАЧ

*В.П. Карп, Б.А. Саруханов*

Процесс приобретения знаний специалистом предметной области рассматривается с позиций нелинейной динамики. Выделены этапы и типы приобретаемых знаний, определяется роль интуитивного мышления в процессе принятия диагностических решений относительно сложных объектов. Рассмотрена роль методов интеллектуального анализа данных в процессе формирования новых знаний. Приводится результат использования разработанных алгоритмов «семантической интерпретации компьютерной базы знаний» на примере конкретной медицинской задачи.

*Ключевые слова:* Процесс приобретения знаний, интуитивные знания, интеллектуальный анализ данных, алгоритмы интерпретации компьютерных знаний, медицина.

### Введение

Процесс приобретения профессиональных знаний специалистом (в любой конкретной области) рассматривается в настоящей работе в аспекте необходимости принятия ответственного решения в сложных для него ситуациях. В данном контексте «сложность ситуации» может оцениваться с различных позиций: отсутствуют априорные сведения об объектах исследования, задача не имеет аналогов, для исследования задано многопараметрическое описание требующей разрешения ситуации, сведения сформулированы не четко, «зашумлены», имеются пропуски и т.п.

Обстоятельств, усугубляющих сложность ситуации, может быть сколь угодно много, но практически все они определяют подобные ситуации как класс слабо формализованных задач и порождают необходимость расширения (приобретения новых) профессиональных знаний, в том числе с использованием современных вычислительных возможностей.

Широкое использование методов искусственного интеллекта позволило достичь значительных результатов при решении различных классов слабо формализованных задач. Особую важность эти методы приобретают при необходимости решать задачи, связанные с исследованием сложно организованных объектов в ситуациях с

неполной и/или зашумленной информацией. Этот круг задач оказался достаточно «популярным» в различных предметных областях и определился как класс задач «распознавания образов».

Технология обучения распознаванию образов на примерах («машинное обучение») оказалась весьма эффективной во многих прикладных областях. Однако трудности, с которыми встречались разработчики при внедрении программ (реализующих решение проблемы), все возрастали. Одна из основных трудностей состояла в том, что большинство правильных решений не воспринималось специалистами. Даже в тех случаях, когда компьютер находил правильное решение, разочаровывал путь решения: промежуточные ходы были «нечеловеческими». При анализе этого феномена выяснилось, что человеком решение принимается как обоснованное и истинное не тогда, когда оно удовлетворяет всем формальным критериям истинности (получено в соответствии с правилами логики), а когда оно правдоподобно, то есть удовлетворяет общим представлениям специалиста и получено в соответствии со сложившимися традициями, принятыми в данной профессии [1].

Возникло новое осмысление проблемы – интеллектуальность должна определяться не только целью, но и способом ее достижения – результат исследования должен быть интерпретирован и представлен в профессиональных терминах специалиста. В настоящей работе на примере решения конкретной медицинской задачи рассматривается принцип организации и технология исследований, предназначенные для проведения интеллектуального анализа данных с целью расширения профессиональных знаний специалиста за счет формализации его интуитивных знаний и интерпретации полученных «машинных» результатов.

### **Исходные позиции. Проблемы расширения знаний специалиста/ эксперта**

К началу своей профессиональной деятельности *специалист* владеет некоторым, не всегда определенным, запасом знаний. Обычно большую долю его знаний на этот момент составляют так называемые *декларативные знания*, приобретенные им в процессе обучения (лекции, учебники, семинары и др.); такие знания считаются условно формализованными, ввиду того что содержат, в основном, общие известные и устоявшиеся понятия специалистов-предшественников, коллег и др., принятые в этой предметной области. Меньшую долю «багажа» составляет пока собственный *накопленный опыт* – результат самостоятельного решения задач в своей профессиональной области. Это, пожалуй, самый ценный, но и самый дорогой способ приобретения (расширения) знаний, так как результат испытания (решения задачи) может не оправдать ожидания (и нанести вред). Тем не менее и при положительном исходе, и при отрицательном формируются (устанавливаются) и запоминаются связи между набором значений параметров эксперимента, оказанным воздействием (фактором влияния) на объект и собственно результатом воздействия. Если удастся обнаружить и зафиксировать эти связи (при том и другом исходе), то можно считать, что эти связи – накопленный опыт – вполне формализованы для множества аналогичных задач. Если же в процессе решения задачи зафиксировать связи не удастся, то считается, что в дальнейшем специалист будет использовать свои *интуитивные знания*. Обычно они не формализованы, но каким-то образом могут оказаться полезными в аналогичных случаях. В процессе накопления профессионализма все большую долю будут составлять «опытные знания» за счет частичной локальной формализации интуитивных знаний.

Реальный переход интуитивных знаний специалиста в «опытные» происходит на основе сопоставления им положительных и отрицательных результатов проведенных экспериментов. При этом основной акцент анализа делается на выявлении и осмыслении неочевидных, на первый взгляд, взаимосвязей между исходными характеристиками объектов, факторами исследования и знаком самого результата (плюс или минус) эксперимента (опыта). Процесс этого перехода, с одной стороны – длительный и зачастую возвратный, так как требуется провести не один-два опыта, чтобы сформулировать достоверные заключения, а с другой – достаточно трудоемкий в силу сложности объектов исследования и самой процедуры анализа.

Очевидно, что перспектива скорейшего решения задачи формализации интуитивных знаний может оказать существенную поддержку и в решении практических задач, и в решении теоретической проблемы, которая до настоящего времени остается одной из важных задач моделирования мышления человека.

### **Исходные позиции. Знания эксперта и возникающие проблемы**

Концепция технологии расширения знаний эксперта с использованием современных методов интеллектуального анализа данных (МИАД), рассмотрена на примере решения медицинских задач.

Медицина представляет собой слабо структурированную область знания, что создает серьезные трудности при построении систем поддержки принятия решений. В то же время, в практической деятельности врач выстраивает последовательность умозаключений, опирающихся на представления о связи наблюдаемых у больного признаков с определенным диагнозом. Быстрота принятия решения зависит, по мнению Б.А. Кобринского [2], как от квалификации и диагностического «чутья» врача, так и от особенности проявления заболевания у конкретного больного. Процесс принятия решения опирается, в первую очередь, на результат диагностики. При этом врач в неявной форме использует аппарат аргументации для подтверждения одной из гипотез – одного или, в редких случаях, сочетания двух диагнозов. Такой аппарат аргументации представляет собой неявные (интуитивные) профессиональные знания врача, накопленные им в результате практической деятельности (опыта). В то же время могут обнаружиться неальтернативные ассоциирующие признаки, обладающие почти одинаковой дифференциальной значимостью, учет которых практически никак не влияет на решение вопроса об отклонении диагностической гипотезы. Подобные ситуации порождают «конфликт» между возникшей сложной проблемой, требующей разрешения, и отсутствием необходимых аргументов для принятия диагностического решения.

По результатам проведенного нами анализа причин возникновения подобных «конфликтов» и способов их разрешения методами искусственного интеллекта, была разработана технология формирования новых знаний в проблеме формализации интуитивных знаний специалиста, представленная на рис. 1.

Следует заметить, что процесс расширения знаний специалиста является динамическим и носит явно нелинейный характер – он определяется, с одной стороны, остротой необходимости решения возникшей задачи, а с другой – наличием возможностей (способов) приобретать новые знания.

К моменту обсуждения конкретной проблемы эксперт – специалист в конкретной предметной области – имеет определенный опыт относительно возникающих си-

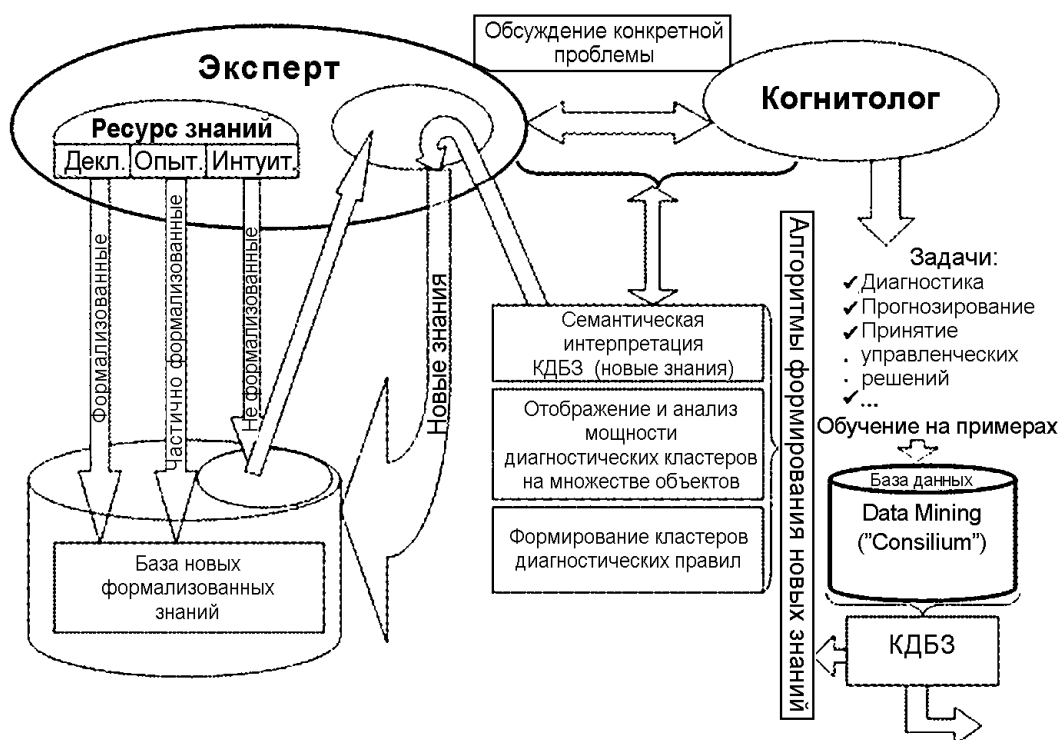


Рис. 1. Концепция расширения знаний эксперта

туаций. Одни ситуации он решает за счет так называемых декларативных знаний – сознательно приобретенных из литературных источников, лекций, конференций и др. Особенность этой категории знаний заключается в том, что они несут в себе сведения о характерных или, наоборот, абсолютно редких (и поэтому очень интересных) прецедентах, представленных другими специалистами. Такие знания могут быть использованы в будущем, но не обеспечивают необходимую надежность «переноса рекомендаций» в конкретном, состоявшемся случае. Другая категория знаний – собственный опыт эксперта, накопленный им за время решения практических задач. Он использует свой навык, зачастую не формулируя аргументацию принятого решения. Сложности начинаются тогда, когда возникают сомнения в выборе решения, и требуется «поддержка аргументацией» выбора одного или другого решения. Такие знания эксперта принято называть «интуитивными».

При возникновении подобных проблем перспективным является использование современных методов интеллектуального анализа данных, представленных в различных пакетах прикладных программ (Data Mining), предназначенных для многостороннего компьютерного исследования возникающих проблем [3–5, 8, 13, 14]. Основная цель таких методов, с одной стороны – обеспечивать корректность анализа на всех этапах исследования, с другой – предоставлять специалисту-исследователю (эксперту в конкретной предметной области) возможность контролировать процесс формирования знаний, используя промежуточные результаты, менять параметры анализа и получать конечные результаты в виде, удобном для интерпретации и т.п. При-

нятые «разумные», с точки зрения специалиста, компьютерные решения конкретной проблемы (решающие правила) получили название «база знаний» (БЗ) или «компьютерная диагностическая база знаний» (КДБЗ).

На рис. 1 отображен весь путь решения проблемы, возникшей у специалиста: от обсуждения содержания самой проблемы, постановки задач, способов решения, и до способов представления результатов анализа.

Рассматриваемая концепция предусматривает два основных этапа решения.

1. Формирование качественной (адекватной решаемой задаче) компьютерной диагностической базы знаний. Принципы и методы классификации изложены в [5].
2. Алгоритмическое обеспечение семантической (смысловой) интерпретации построенных баз знаний в профессиональных терминах специалиста.

Следует заметить, что первый этап включает решение проблемы классификации, а второй – разработку дополнительных алгоритмов для выявления и учета связей между фрагментами диагностических правил, визуализации выявленных в КДБЗ кластеров связей, обеспечивающих семантическую интерпретацию обнаруженных знаний. В данной работе этой части разработок уделяется основное внимание.

В дальнейшем полученные новые знания, сформулированные на профессиональном языке, формализованные и аргументированные, могут быть использованы в различных вариантах их тиражирования: в виде декларативных знаний, в различных консультационно-диагностических системах, экспертных системах, в системах поддержки принятия решений [5, 6, 13, 15, 16] и др.

Суть предлагаемой концепции заключается в следующем. В результате компьютерного анализа формируется так называемая компьютерная диагностическая база знаний, содержащая списки выявленных связей – диагностических правил (ДП), характерных для каждого из сравниваемых классов. Но несмотря на значительные реальные результаты при решении различных классов слабо формализованных задач, в плане расширения профессиональных знаний специалиста/эксперта построенные КДБЗ до последнего времени не использовались полностью. Это объясняется тем, что компьютерные реализации, в основном, были ориентированы на достижение эффективности распознавания, и перед ними не ставилась задача «расшифровки» полученных диагностических правил на «языке» пользователя.

Возрастание потребностей пользователей в плане понимания полученных ДП определило необходимость разработки методов дополнительного анализа построенных компьютерных баз знаний и алгоритмов, обеспечивающих максимальную интерпретацию совокупности выявленных закономерностей.

Таким образом, основными задачами данного исследования явились: разработка методов дополнительного анализа построенных диагностических баз знаний и компьютерная реализация алгоритмов, обеспечивающих интерпретацию совокупности выявленных закономерностей на профессиональном языке пользователя, с целью обобщения семантической картины проблемной области.

Эффективность подхода продемонстрирована на примере решения конкретной медицинской проблемы [9–12].

## **Реализация технологии формализации интуитивных знаний**

Технология формализации интуитивных знаний специалиста реализована на базе модульной системы поддержки принятия решений «Consilium» [13], обеспечивающей проведение «машинного обучения на примерах», включая формирование адекватного языка описания объектов и формирование проблемно-ориентированной базы данных, преобразование исходного пространства признаков в «условно оптимальное» для поставленной задачи, выявление закономерностей, характеризующих каждый из заданных классов сравнения, построение и оценка диагностических правил, содержащихся в КДБЗ. В основе классификации используются методы поиска характерных закономерностей, часто называемых *методами перебора конъюнкций* [5,7,13,14]. Логическая схема этой группы алгоритмов основывается на эвристических соображениях о деятельности человека при распознавании образов (идея предложена М.М. Бонгардом).

В работе [5] детально изложены основные принципы и этапы решения различных классов слабо формализованных задач; определены модули, допускающие максимальную алгоритмизацию анализа данных; предложены алгоритмы, повышающие эффективность функционирования некоторых модулей; рассмотрены критерии, определяющие качество построенного решающего правила и расчет его диагностической ценности; определен горизонт практического использования КДБЗ. Там же приведена таблица сравнения результатов диагностики различными методами классификации.

В настоящее время для расширения возможностей технологии формализации интуитивных знаний разработаны дополнительные эвристические алгоритмы: «Alost – формализация интуитивных знаний эксперта» [8] и «САБЗ – семантический анализ базы знаний» [9–11,15,16], позволяющие «перевести» (интерпретировать) компьютерные закономерности в профессиональные термины специалиста. Ниже приводится описание алгоритмов САБЗ.

### **Алгоритмы семантического анализа диагностических баз знаний**

Интерпретация экспертом результатов работы алгоритмов Data Mining является важным этапом в технологии создания баз знаний по конкретной проблеме (в частности, для решения задачи классификации). Успех на данном шаге в большой степени зависит от полноты картины, предоставленной для анализа. В процессе решения задач классификации объектов (практически любой предметной области) возникает проблема учета особенностей их описания, особенно это относится к сложно организованным объектам. Это связано с тем, что сложные объекты, как правило, характеризуются большим количеством качественных (неизмеряемых) показателей (признаков) или, например, наличием пропущенных данных, порожденных невозможностью непосредственного измерения интересующего показателя. Это вынуждает исследователей дополнять пространство основных признаков косвенными характеристиками в расчёте на то, что при отсутствии части сведений о диагностируемом объекте можно будет использовать имеющиеся косвенные признаки как для выявле-

ния общих закономерностей, характерных для каждого класса состояний, так и для диагностики конкретного объекта. Однако ДП, построенные на таком смешанном пространстве, могут оказаться статистически или содержательно коррелированными между собой, что может исказить результаты и диагностики, и осмысления проблемной области в целом.

Выходом из данной ситуации является уточнение и расширение представлений эксперта об обобщённой картине проблемной области, формируемой набором ДП. В этом случае специалист может интерпретировать не только каждое правило в отдельности, но и оценить всю совокупность правил БЗ, построенных для каждого из сравниваемых классов (А – «Неблагополучные», В – «Благополучные»). Разработанный алгоритм САБЗ [16] предназначен для формирования у специалиста обобщённого представления о структуре проблемной области. На рис. 2 представлена схема алгоритма. Идея его основывается на гипотезе о том, что симптом, образующий наибольшее количество связей с другими симптомами, может выступать в качестве базового, образующего кластер смыслового объединения этих симптомов.

Работа алгоритма начинается со сбора сведений о том, каков состав каждого правила в КДБЗ для каждого из классов (А и В); у какого количества объектов оно проявляется, и у каких именно. Данная информация позволяет сформировать картину взаимосвязей между самими ДП внутри одного класса. Этот аспект весьма важен для данной ситуации, поскольку представление базы знаний в виде наборов продукционных правил (а именно такая форма представления БЗ подвергается анализу) не позволяет явно оценить степень взаимосвязи между двумя отдельными конъюнкциями.

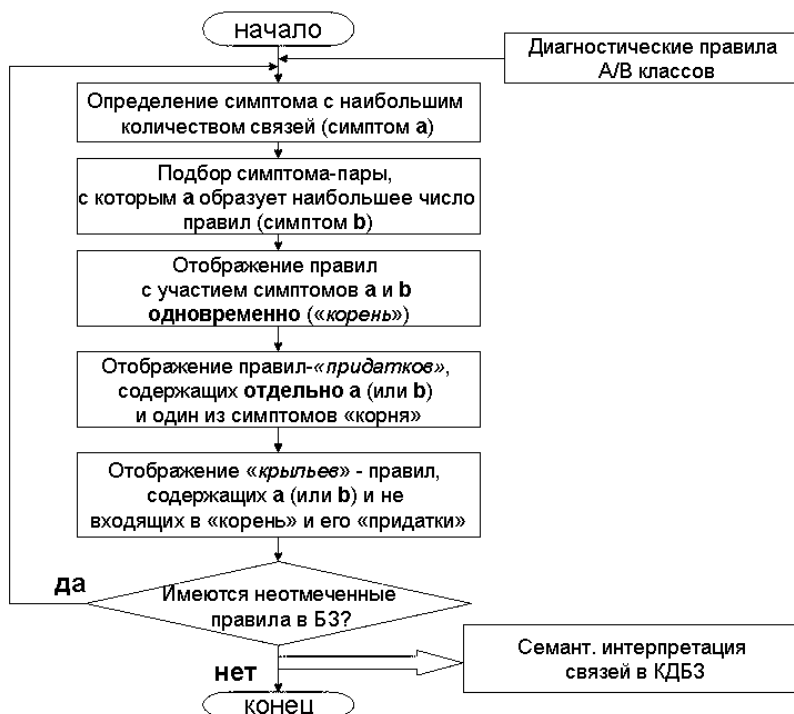


Рис. 2. Схема алгоритма семантического анализа базы знаний

Применение алгоритма САБЗ позволяет выявлять группы ДП БЗ, близкие по смысловому составу; такие группы названы кластерами. Семантический анализ на основе кластеров позволяет получить более полную картину ситуации в каждом конкретном сегменте проблемной области. Это, в свою очередь, облегчает интерпретацию экспертом результатов диагностики.

На рис. 3 представлен фрагмент (доминирующий кластер) выявленных алгоритмом связей между ДП КДБЗ, построенных для одного из классов сравнения (для класса А).

Рассмотрим структуру кластера. Симптом, имеющий наибольшее количество связей с другими симптомами базы знаний («базовый») будет располагаться в основе схемы связей в паре с симптомом, с которым он образует наибольшее количество правил. В результате на схеме отображены два «базовых» симптома (55 и 59), совместно образующие подкластер 1, а также два других подкластера (2 и 3), каждый из которых образуется своим «ведущим» симптомом.

В первый подкластер входят ДП, в которых участвуют оба базовых симптома (такие правила называются «корневыми»), а также «смежные» с ними правила. Под смежными в данном случае понимаются ДП, имеющие в своём составе один из базовых симптомов и один из симптомов корневых правил (например, 59-289).

Во второй подкластер входят оставшиеся не отмеченными на схеме ДП из базы знаний, содержащие первый базовый симптом (например, 55-94-77).

В третий – содержащие второй базовый симптом (например, 59-126-77). Эти два подкластера (второй и третий) называются «крыльями». Каждый кластер может содержать до трех подкластеров.

«Доминирующий» кластер (обычно первый), построенный на основе наиболее «активного» симптома в базе знаний, обладает важным свойством: он всегда

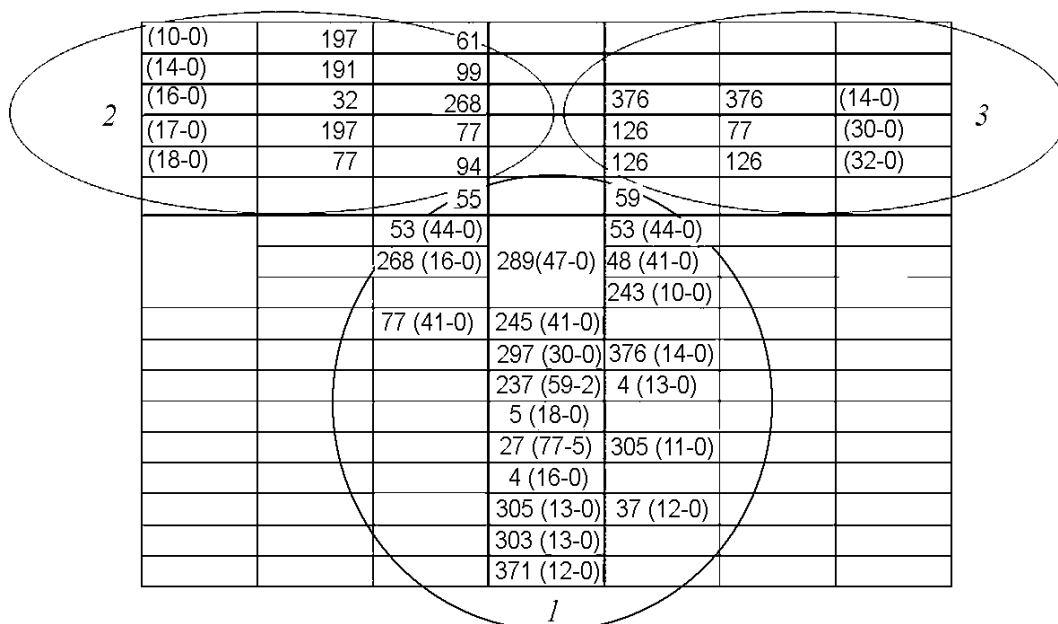


Рис. 3. Доминирующий кластер диагностических правил для класса А («Неблагополучные»)



охватывает большее количество объектов исследуемого класса, чем все последующие кластеры. Поэтому базовые элементы этого кластера могут оказаться наиболее значимыми (информативными) при принятии решения по принадлежности объекта к классу.

На рисунке построенного кластера в ячейках рядом с номерами симптомов, участвующих в конъюнкциях, в скобках указано количество объектов, у которых было обнаружено данное правило (через дефис указаны количества объектов класса А – «Неблагополучные» и класса Б – «Благополучные», соответственно). Над базовыми симптомами изображены «крылья» – подкластеры, в составе которых есть только один из двух базовых симптомов.

Последующие кластеры могут оказаться частично или полностью в зоне охвата доминирующего кластера или охватывать область, не смежную с ним. Это может означать, что в материале исследования присутствуют объекты-прецеденты, то есть редко встречающиеся случаи.

Анализ и интерпретация содержания всех остальных выявленных кластеров проводится аналогичным образом.

#### **Интерпретация кластеров связей компьютерной диагностической базы знаний по конкретной медицинской задаче**

Медицинская (прикладная) цель исследования была сформулирована следующим образом: «Приобретение новых объективных знаний относительно оценки состояния новорожденных, родившихся от матерей, перенесших сифилис» (Задача «ДВС»).

Требовалось:

- провести сравнительный анализ данных неблагоприятного и благополучного исхода беременности и определить набор показателей, характерных для этой категории пациентов;
- сформировать диагностические правила (ДП) оценки состояния новорожденных при неблагоприятном и благополучном исходе беременности;
- выявить группы риска относительно основных факторов заболевания ВС;
- сравнить группы новорожденных при скрытом раннем сифилисе у матери;
- определить диагностические критерии оценки состояния новорожденных при отсутствии у них клинических проявлений ВС.

Исходный материал исследования составили данные по 1363 случаям рождения детей от матерей, серопозитивных по сифилису, полученные в результате проведенного в 2003–2004 годах эпидемиологического исследования в семи регионах РФ. В исследование вошли 315 анкет: из них – все подтвержденные случаи «неблагополучного» исхода (кл. А – 165) и с «благополучным» исходом (кл. В – 150).

С использованием модульной системы «Consilium» [5,13], основанной на методе перебора конъюнкций, были построены диагностические правила для каждого из заданных классов исхода (А и В). Сформированная КДБЗ состоит из двух списков диагностических правил, представленных в виде конъюнкций номеров симптомов (симптом – одно из возможных значений признака).

В целом, КДБЗ представлена в виде продукционной модели. В диагностических списках, допущенных к голосованию, оказалось порядка 50–60 правил (для каждого класса). На основе КДБЗ разработана консультативно-диагностическая интеллектуальная система для диагностики новорожденных (ИС-ДВС) [15].

В табл. 1 приведен фрагмент базы знаний – диагностические правила, сформированные по классу А. В первом столбце данной таблицы приведены порядковые номера некоторых выбранных для иллюстрации ДП. Во втором – номера тех симптомов, которые входят в конкретное ДП. Далее два столбца, отражающие количество объектов (из класса А и из класса В), имеющих данное ДП. Затем – рассчитанное значение диагностической информативности ДП. В крайнем правом столбце приведен содержательный смысл каждого ДП. Аналогично представляются ДП класса В.

При диагностическом голосовании для конкретного пациента каждое ДП участвует независимо от других, что вполне приемлемо, если эти ДП независимы между собой, в противном случае требуется учитывать это обстоятельство.

Из таблицы видно, что при большом объеме ДП очевидным преимуществом является представление базы диагностических правил в виде кластеров.

На рис. 4 изображены подкластеры 1, 2, 3 – фрагменты кластера, представленного ранее на рис. 3, и их интерпретация специалистами.

Таблица 1

Пример диагностических правил для класса А

№ ДП (п/п)	Номера симптомов входящих в ДП			Количество объектов из класса		Вес ДП	Предметное описание диагностической закономерности
				А	В		
1	55	59	289	47	0	3.384	Отсутствие дородового лечения матери, приведшее к положительным н/т у новорожденного
15	126	77	48	30	0	2.954	Диагноз сифилиса у матери, установленный на сроках беременности $\geq 33$ недель, явившийся причиной клинических проявлений ВС у новорожденного
24	77	77	197	20	0	2.574	Отсутствие профилактического лечения у матери, когда во втором триместре беременности показатель индекса н/т находился в пределах значений 0.75–1.0 (3+)
34	4	55	59	16	0	2.370	Отсутствие дородового лечения у матерей в возрастной группе < 19 лет
50	61	61	197	10	0	1.955	Неадекватное лечение матери (непенициллиновая терапия, либо пенициллиновая терапия, законченная за 30 дней до родов), индекс н/т во втором триместре беременности которой находился в пределах значений 0.75–1.0 (3+)
61	237	55	36	15	1	1.801	Отсутствие специфического лечения вторичного сифилиса у матери до беременности при сохранении степени позитивности н/т к моменту родов, равной 4+

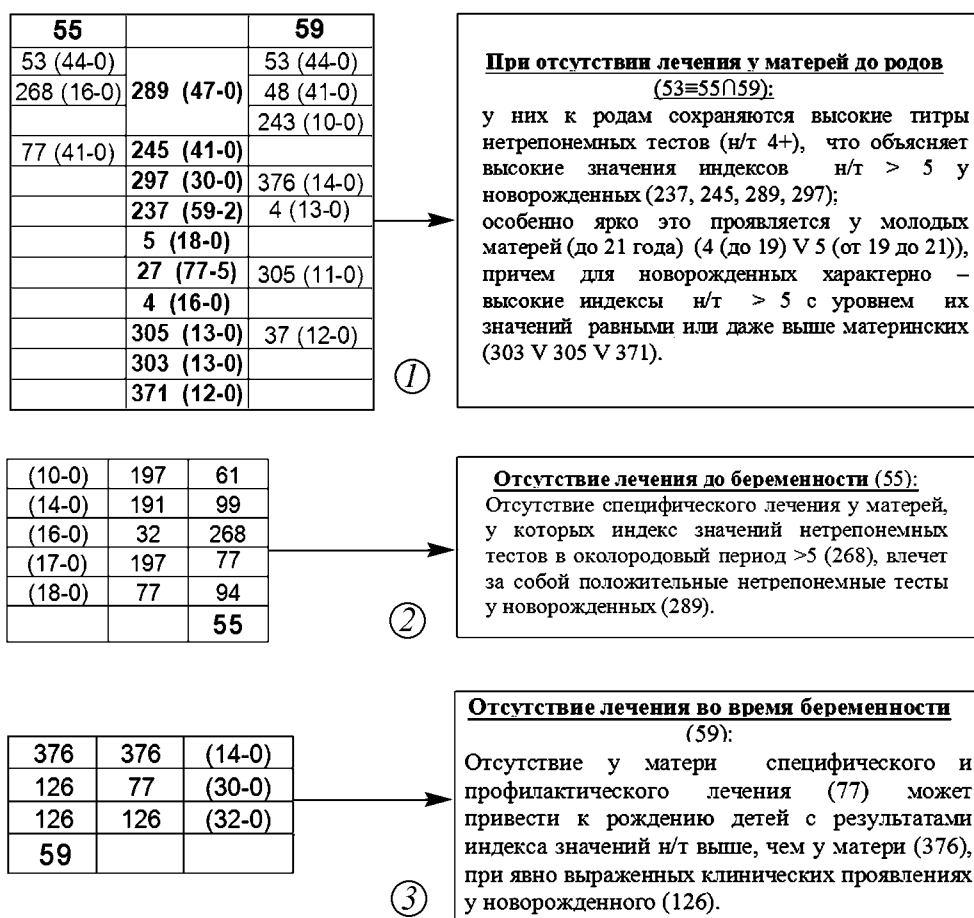


Рис. 4. Подкластеры 1, 2, 2 доминирующего кластера класса А

Интерпретация обобщенной картины связей между участвующими ДП становится очевидной и облегчает осмысление основных (типичных, характерных и т.п.) закономерностей проблемной области.

Семантическая интерпретация данной структуры показала, что наиболее «активным» симптомом в правилах класса «Неблагополучные» оказался симптом под номером 55 – «объект не лечился в первой половине наблюдений». Больше всего конъюнкций с ним образовал симптом 59 – «объект не лечился во второй половине наблюдений». По сути, все правила корневого подкластера, включая смежные боковые, подразумевают, что объект наблюдений не проходил лечения вовсе (симптом 53 ≡ 55 ∩ 59). Диагностические правила БЗ, содержащие данные симптомы, существенно чаще встречаются в КДБЗ класса «Неблагополучные».

Форма представления содержательной структуры проблемной области в виде кластеров позволяет обнаружить (увидеть) дополнительные связи между симптомами, участвующими в разных диагностических правилах БЗ, но оказавшимися в составе одного кластера. Это облегчает задачу специалисту, ведущему построение обобщенной картины проблемной области по совокупности анализа всех кластеров, выявленных алгоритмом.

## Отображение мощности диагностических кластеров на множестве объектов исследования

Параллельно с построением кластеров система определяет и те объекты, которые оказываются «охваченными» формируемыми кластерами. Для этого регистрируется информация о том, какие правила были обнаружены у каждого объекта.

Для каждого кластера указывается список номеров объектов в БЗ, охватываемых им, а также общее количество охваченных объектов каждого из рассматриваемых классов. Это позволяет оценить диагностический вес кластера, например, с помощью функции правдоподобия [5].

Алгоритм отображения мощности диагностических кластеров предназначен для ответа на очень важный для специалиста вопрос – каков состав объектов, охваченных кластером в целом и его подкластерами.

Если кластер охватывает значительное количество объектов одного класса (и почти не «захватывает» объекты другого класса), то он интерпретируется экспертом как значимый, характерный (типичный) для данного класса диагноза, даже если он не был известен ранее.

Если кластер охватывает лишь небольшую группу объектов в одном классе диагнозов, то, скорее всего, он отражает подмножество объектов со специфическими характеристиками (прецеденты).

Если же кластер охватывает небольшую группу объектов как в одном, так и в другом классе, – такой кластер не очень полезен с точки зрения решения конкретной задачи диагностики, но может нести ценную информацию о наличии объектов в переходном состоянии или образующих подкласс на данном множестве объектов по каким-то иным причинам.

В табл. 2 представлен пример расчета параметров мощности доминирующего кластера класса «Неблагополучные» (представленного на рис. 3).

Для каждого подкластера (1, 2, 3) и для кластера в целом (1V2V3) приведены расчетные величины и количества объектов, входящих в каждый из них, а также их доля относительно общего количества объектов в соответствующем классе.

Оценка индивидуального и совместного вклада подкластеров конкретного кластера учитывается алгоритмом в два этапа. Индивидуальным вкладом подкластера считается количество объектов, имеющее конъюнкции этого подкластера, за исключением тех объектов, которые могут быть узнаны двумя другими подкластерами этого же кластера. Так, например, потенциал подкластера 1 составляет (97-5) объектов, а цена его вклада при наличии двух других подкластеров составляет (49-5), так как остальные 48 объектов (15+16+17) могут быть узнанными двумя другими подкластерами (2 или 3). Из полученного анализа можно сделать вывод, что кластер в целом характерен для 72% объектов класса «Неблагополучные», и есть подозрение, что 3% объектов другого класса – «Благополучные» – имеют сходные характеристики с объектами подкластера 1, входящего в этот кластер.

Таблица 2

Мощность подкластеров и кластера

Подкластер	1	2	3	Мощность кластера (1V2V3)
Мощность подкластера (кол-во объектов)	97-5	53-0	34-0	119-5
Мощность подкластера (%)	59-3	32-0	21-0	72-3

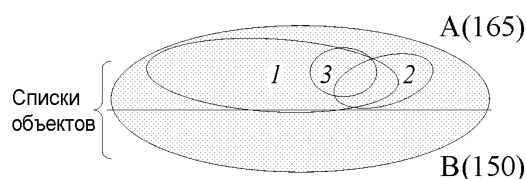


Рис. 5. Оценка мощности кластеров на множестве объектов исследования (кл.А)

На рис. 5 приведен пример графического отображения результатов работы алгоритма (см. табл. 2). Изображены доминирующий кластер и три его подкластера, охватывающие достаточно большую часть объектов – 72%, входящих в класс А, и включающие 3% объектов класса В.

Таким образом, на основе оценки мощности кластера можно сформулировать семантическую оценку степени отражения им обобщенной картины проблемной области заданного класса.

### Заключение

Рассмотрены проблемы, возникающие у специалиста в процессе приобретения и расширения своих профессиональных знаний. Определена роль интуитивных знаний в принятии решений относительно сложно организованных объектов. Сформулированы условия и факторы, обеспечивающие мотивацию расширения своих профессиональных знаний.

Показано, что процесс расширения знаний специалиста является динамическим и носит явно нелинейный характер – он определяется, с одной стороны, острой необходимостью решения возникшей проблемы, а с другой – наличием возможностей (способов) приобретения новых знаний.

Рассмотрены возможности и перспективы использования современных методов интеллектуального анализа данных.

Разработана технология и представлена концепция расширения профессиональных знаний эксперта. Технология обеспечивает выполнение всех этапов решения проблемы, возникшей у специалиста: от обсуждения самой проблемы, постановки задачи, способов решения и способов представления результатов анализа.

В рамках предложенной технологии разработан программный комплекс алгоритмов САБЗ, предназначенный для формирования у специалиста обобщенного представления о структуре проблемной области.

Эффективность использования методов интеллектуального анализа данных и разработанных алгоритмов семантического анализа базы знаний продемонстрирована на реальной медицинской задаче.

### Список используемых сокращений

БЗ – база знаний

ВС – шифр медицинского диагноза

ДВС – шифр конкретной медицинской задачи

ДП – диагностическое правило

ИС-ДВС – интеллектуальная система диагностики ВС

КДБЗ – компьютерная диагностическая база знаний

САБЗ – семантический анализ базы знаний

## Библиографический список

1. *Губерман Ш.А.* Искусственный интеллект-88 и «Продуктивное мышление» М. Вертгеймера (1945; рус. перевод М.: Прогресс, 1987) // Вопросы психологии. 1990. № 1. Тематические сообщения. С. 100.
2. *Кобринский Б.А.* Логика аргументации в принятии решений в медицине // НТИ, сер. 2, 2001, № 9. С. 1.
3. *Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И.* Методы и модели анализа данных: OLAP и Data Mining. СПб.: БХВ-Петербург, 2004. 336.
4. *Дюк В., Самойленко А.* Data Mining. СПб.: Питер, 2001.
5. *Карп В.П.* Интеллектуальный анализ данных в проблеме построения решающих правил классификации //Новости искусственного интеллекта. 2006. № 2. С. 57.
6. *Чернавский Д.С., Карп В.П., Родитат И.В. и др.* Распознавание. Аутодиагностика. Мышление. Синергетика и наука о человеке. М.: ИПРЖР, 2004.
7. *Бонгард М.М.* Проблемы узнавания. М.: Наука, 1967. 320.
8. *Карп В.П., Огурчикова А.Л.* Интеллектуальная система структурного анализа групп сравнения сложноорганизованных объектов // Сб. научных трудов 59-й научно-технической конференции МИРЭА. М., 2010, Ч. 1. С. 14.
9. *Карп В.П., Саруханов Б.А., Сафарова Э.А.* Алгоритм обнаружения и учета скрытых связей в диагностических базах знаний по конкретной проблеме // Сб. научных трудов 59-й научно-технической конференции МИРЭА. М., 2010, Ч. 1. С. 18.
10. *Карп В.П., Саруханов Б.А., Сафарова Э.А.* Алгоритм семантического анализа диагностической базы знаний в системах поддержки принятия решений // Материалы XI ежегодной специализированной конференции и выставки «Информационные технологии в медицине». М.: «Консэф», 2010. С. 154.
11. *Карп В.П., Саруханов Б.А.* Современные интеллектуальные технологии анализа данных в развитии диагностических баз знаний // Сборник научных тезисов XVIII Международ. конференции «Математика. Компьютер. Образование». Пущино, 2011. С.100
12. *Карп В.П.* Интеллектуальный анализ данных в проблеме формализации интуитивных знаний эксперта // Сб. трудов конференции «Нелинейная динамика в когнитивных исследованиях» 2011: Ин-т приклад. физики. Нижний Новгород: ИПФ РАН, 2011. С. 82.
13. *Карп В.П.* Программа для ЭВМ для построения решающих правил классификации (Консилиум-1) // Всероссийский фонд алгоритмов и программ (РосАПО). М., 1994. № 940206. 28 с.
14. *Карп В.П., Чернавский Д.С., Никитин А.П.* Способ определения кинетических параметров электрических потенциалов точек акупунктуры при диагностике организма и устройство для его осуществления. Государственный реестр изобретений, Патент № 2093134 от 20.10.97.
15. *Карп В.П., Саруханов Б.А.* Интеллектуальная система диагностики заболеваний

новорожденных от инфицированных матерей. Патент РФ на полезную модель Государственный реестр полезных моделей РФ № 108675 от 20.09.2011.

16. *Карп В.П., Саруханов Б.А.* Система анализа базы знаний. Авторское Свидетельство Гос. Рег. программы на ЭВМ № 2011617361. Реестр программ для ЭВМ 21.09.2011.

*Физический институт  
им. П.Н. Лебедева РАН  
Московский государственный  
институт радиотехники,  
электроники и автоматики*

*Поступила в редакцию 12.07.2011  
После доработки 14.11.2011*

### **NONLINEAR DYNAMICS OF AN EXPERT'S THINKING AND ALGORITHMS OF HIS PROFESSIONAL KNOWLEDGE EXPANSION BY EXAMPLE OF MEDICAL PROBLEMS**

*V.P. Karp, B.A. Sarookhanov*

The process of a problem area experts knowledge acquisition is considered from a nonlinear dynamics positions. Stages and types of acquired knowledge are marked out; the role of an intuitive thinking in decision process concerning complex objects is defined. The role of intelligent data analysis methods in new knowledge forming processes is defined. The result of a developed «Semantic computer knowledge base analysis» algorithm is given on an example of the concrete medical problem.

*Keyword:* Knowledge acquiring process, intuitive knowledge, intelligent data analysis, computer knowledge interpretation algorithms, medicine.

*Авторские сведения Карп В.П. см. на стр. 20.*



*Саруханов Баграт Артурович* – родился в Москве (1985), окончил Московский государственный институт радиотехники, электроники и автоматики (технический университет, 2002). Является аспирантом кафедры информационных систем МИРЭА с 2009 года по направлению «Системный анализ, управление и обработка информации». Автор 8 научных статей по указанному направлению.

119270 Москва, пр-т Вернадского, 78  
Московский государственный институт радиотехники,  
электроники и автоматики  
E-mail: [allerit@yandex.ru](mailto:allerit@yandex.ru)