

и уравнения для определения постоянных Θ_1, Θ_2 :

$$\cos \Theta_1 + \cos \Theta_2 = -A_0 / ma;$$

где $0 \leq \frac{A_0}{ma} \leq 1$; $\sin \Theta_1 + \sin \Theta_2 = 0$. (7)

Из (6) следует, что центр масс шатуна, ползуна и массы m_k находится в точке D шарнирного соединения шатуна с кривошипом. Это одно из условий уравновешивания шарнирно-ползунного механизма, которое приводится в [3].

Система ур. (7) имеет решение, соответствующее реальному положению маятников:

$$\Theta_2 = -\Theta_1; \quad \Theta_1 = \arccos\left(-\frac{A_0}{2ma}\right). \quad (8)$$

Расположение маятников, соответствующее ур. (8), показано на рис. 2.

Следовательно ур. (7, 8) является условием отсутствия движения платформы, т.е. условием уравновешивания механизма.

Ур. (1, 2) при выполнении (6) совпадают с уравнениями движения ротора с шарами [2]. Отсюда ус-

ловие устойчивости стационарного движения (4, 5, 8), т.е. условие статического уравновешивания механизма при удовлетворении его параметров соотношениям (6) и (9), по аналогии с [2], запишем в виде:

$$\omega_\xi < \Omega < \omega_p; \quad \omega_\eta < \Omega,$$

где

$$\omega_\xi = \frac{C_\xi}{M}; \quad \omega_\eta = \frac{C_\eta}{M}; \quad \omega_p^2 = (\omega_\xi^2 + \omega_\eta^2)/2; \quad \omega_\xi < \omega_\eta.$$

Рассмотренный способ уравновешивания кривошипно-ползунного механизма с помощью маятников более предпочтителен, чем рассмотренный в работе [3], где для уравновешивания механизма используют противовесы на кривошипе и шатуне так, что центр масс остается неподвижным. Преимущество изложенного способа заключается в том, что он осуществляется автоматически за счет синхронизации движения маятника и кривошипа.

Авторы отмечают, что статья является развитием работ В.П. Нестеренко, внесшего большой вклад в разработку принципов автоматической балансировки роторов.

СПИСОК ЛИТЕРАТУРЫ

1. Блехман И.И. Синхронизация динамических систем. — М.: Наука, 1971. — 796 с.
2. Нестеренко В.П. Автоматическая балансировка роторов приборов и машин со многими степенями свободы. — Томск: Изд-во Томского ун-та, 1985. — 85 с.
3. Щепетильников В.А. Уравновешивание механизмов. — М.: Машиностроение, 1982. — 256 с.

УДК 004.3:681.3

ВЫЯВЛЕНИЕ СКРЫТЫХ ЗАКОНОМЕРНОСТЕЙ В СОЦИАЛЬНО-ПСИХОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

Е.А. Муратова, О.Г. Берестнева

Томский политехнический университет
E-mail: helmu@mail.ru

Рассматриваются методы выявления скрытых закономерностей, позволяющие выявить присущие исследуемой предметной области устойчивые закономерности в виде логических правил, с последующим построением их метаструктуры. Описывается принцип формирования фиктивных бинарных переменных. Построение метаструктуры является весьма существенным для построения баз знаний, требующих ввода понятий, метапонятий и семантических отношений на основе множества фрагментов знаний о предметной области.

Введение

При решении задачи формирования базы знаний для интеллектуальных медицинских (в том числе психологических и психодиагностических) систем используют методы, позволяющие выявить присущие исследуемой предметной области устойчивые закономерности на основе имеющихся данных с привлечением или без привлечения экспертов. Следовательно, результаты решения одной и той же

диагностической задачи разными методами будут в какой-то мере отличаться друг от друга. На наш взгляд, совместное использование полученных решений позволит повысить качество распознавания, классификации и прогнозирования при использовании минимального количества диагностических прецедентов.

Имеющийся опыт работы со специалистами-диагностами (врачами, психологами, психотерапевтами и др.) показал, что математические решения

поставленной задачи воспринимаются легче, если они представлены в виде логических высказываний. Поэтому целью данной работы была разработка технологии конструирования диагностических решений в виде логических правил. Построение логических правил (в дальнейшем также диагностических решений) позволяет воспроизвести процесс принятия диагностического решения экспертом изучаемой предметной области.

Задача построения логических правил не нова [1–7]. Однако, до настоящего времени, остается актуальной задача разработки технологии совместного использования разных диагностических решений, что будет особенно ценным, например, для выборки малой размерности, характерных для социологии, психологии и психодиагностики.

1. Постановка задачи и описание методов решения

Исходным материалом при решении задач анализа данных является набор объектов предметной области, представленный характеризующими его признаками, которые измеряются в шкалах разного типа.

Требуется получить диагностическое решение, представляющее собой конъюнкцию элементарных событий. В качестве элементарных событий рассматриваются отдельные значения какого-либо признака, и, следовательно, логическое правило будет иметь следующий вид: Если *условие1 и условие2 и ... и условиеK* То *исход1*. Любое правило в виде условного суждения Если ... То ... имеет две основные характеристики – точность и полноту. Точность правила – это доля случаев, когда правило подтверждается на всей совокупности данных. Полнота правила – это доля случаев, когда правило подтверждается на объектах, для которых характерен только *исход1*.

Другими словами, для построения логических решающих правил требуется найти подмножество признаков из заданного числа признаков и указать области их значений, при которых качество распознавания обучающей выборки будет не ниже заданного.

Число подмножеств из k признаков, которые надо просмотреть, достаточно велико и может быть оценено по формуле

$$C_l^k = \frac{l!}{k!(l-k)!},$$

где l – общее число признаков.

Для высоких размерностей полный перебор является абсолютно нереальным. Поэтому существуют различные методы, позволяющие избежать полного перебора признаков [1, 8–12]. В данной работе будут рассмотрены следующие методы поиска информативных подмножеств признаков:

- 1) локальной геометрии;
- 2) ограниченного перебора;
- 3) усечения для построения деревьев решений.

Выбор данных методов обусловлен доступностью пакетов, реализующих вышеуказанные методы, а также возможностью в достаточно короткий срок освоить работу с этими пакетами.

Метод локальной геометрии позволяет рассматривать любой объект обучающей выборки как самостоятельный классификатор, обладающий полезными свойствами для решения конкретной диагностической задачи. Получаемая система логических правил, как показано в [7, 13], позволяет улучшить классификацию и прогноз по сравнению с классическими математическими методами (дискриминантный анализ, кластерный анализ и др.). Для реализации данного метода необходимы элементарные навыки работы с электронными таблицами и наличие статистических пакетов, например, Statgraphics for Windows.

Метод ограниченного перебора впервые был предложен в середине 60-х годов XX века М.М. Бонгардом [12]. С тех пор подобные алгоритмы продемонстрировали свою эффективность при решении множества задач из самых различных областей. Эти алгоритмы вычисляют частоты комбинаций простых логических событий в подгруппах данных, а ограничением служит длина таких комбинаций. На основании анализа вычисленных частот делается заключение о полезности той или иной комбинации для распознавания, классификации и прогнозирования. Данный метод реализован в системе WizWhy, которая достаточно удобна при эксплуатации. Максимальное число комбинаций логических высказываний в данной системе равно 6.

Деревья решений являются одним из наиболее популярных подходов к решению задач распознавания, классификации и прогнозирования и представляют собой логическую модель закономерностей исследуемого явления. Они воссоздают иерархическую структуру классифицирующих правил типа "Если ... То ...", имеющую вид дерева. Для того чтобы решить, к какому классу отнести некоторый объект, требуется ответить на вопросы, стоящие в узлах этого дерева, начиная с его корня. Вопросы имеют вид: "Значение признака x больше a ?". Если ответ положительный, осуществляется переход к правому узлу следующего уровня, если отрицательный – то к левому узлу; затем снова следует вопрос, связанный с соответствующим узлом. Популярность подхода связана с наглядностью и понятностью получаемых решений. В данной работе была использована система See5, в которой используется метод усечения, позволяющий дать объективную оценку качества дерева решений.

2. Формирование фиктивных бинарных переменных

Определяя диагноз, специалист сравнивает имеющиеся значения показателей с граничными значениями и, исходя из этого, делает свое заключение. Для того, чтобы восстановить процедуру при-

нятия решения по имеющейся базе данных, целесообразно перевести количественные признаки в серию бинарных признаков. Это позволит сделать получаемые логические правила более "правдоподобными" или "естественными".

Каким же образом осуществить данный переход? Для решения этой задачи, как правило, используется разбиение на интервалы, предлагаемое экспертом. Или диапазон значений признака разбивается на равные по длине подинтервалы, число которых также задается экспертом. Поэтому представляется целесообразным осуществлять преобразование количественных признаков в бинарные с учетом особенностей решаемой задачи, не привлекая экспертов [13, 14]. В системе WizWhy и See5 разбиение на информативные интервалы осуществляется в процессе построения логических правил. Для метода локальной геометрии требуется предварительное выделение информативных интервалов. Рассмотрим этот подход.

За основу неравномерного разбиения был взят метод адаптивного кодирования признаков, предложенный Н.Н. Тарановой [14], который нами был адаптирован для работы с малыми выборками.

Выделение информативных подинтервалов, основано на предположении о том, что признаки, наиболее полно характеризующие различие двух диагностических классов, по своей физической природе таковы, что можно их дискретизировать таким образом, чтобы отношение относительных частот встречаемости объектов обучающих выборок этих двух классов в выделенных подинтервалах могло быть аппроксимировано монотонной или одноэкстремальной функцией [14].

Вышеуказанный модифицированный алгоритм был реализован программно с использованием инструментального средства Delphi фирмы Borland International в среде Windows 95/98 и включен в систему UniData, позволяющей формировать бинарные признаковые пространства как с привлечением эксперта, так и самостоятельно, адаптируя их к особенностям решаемой задачи. В систему UniData включены различные способы преобразования, а именно:

1. Экспертное выделение подинтервалов, где предусмотрено два варианта: разбиение с количеством интервалов, задаваемым экспертом, и разбиение, предлагаемое самим экспертом.
2. Равномерное выделение подинтервалов.
3. Неравномерное выделение подинтервалов на основе модифицированного алгоритма адаптивного кодирования разнотипной информации.
4. Комбинированное выделение подинтервалов, основанное на комбинировании способа равномерного выделения интервалов с последующим объединением нулевых подинтервалов.

Таким образом, использование нескольких способов преобразования количественных признаков в бинарные, позволяет получать различные вариан-

ты диагностических решений и выбирать логические правила, имеющие лучшую степень распознавания и прогнозирования.

3. Конструирование диагностических решений

Построение различных логических моделей позволяет выявить устойчивые закономерности, представляющие собой диагностическое сочетание характеристических признаков, так называемых "синдромов" (например, в медицине "синдром" — это сочетание признаков, характерное для определенного заболевания) [15], и тем самым расширить базу знаний исследуемой предметной области. Формирование расширенного признакового пространства может осуществляться путем добавления к исходному пространству новых признаков, порожденных наборами диагностических сочетаний. Соответственно, уменьшение признакового пространства может быть осуществлено путем замены набора диагностических признаков на их обобщенный ("синдромный") признак.

Ниже будут описаны методы построения таких моделей с использованием как традиционных статистических (Statgraphics), так и специализированных (WizWhy и See5) пакетов.

3.1. Построение логических моделей методом локальной геометрии

Нахождение логических закономерностей в локальном пространстве признаков заключается в комбинированном применении методов многомерного линейного анализа данных и интерактивной графики, которые успешно реализованы в современных статистических пакетах.

Постановка данной задачи не нова и содержится в [7, 16]. Поскольку метод локальной геометрии полностью не автоматизирован, то ниже кратко изложена схема нахождения логических закономерностей в прикладном статистическом пакете Statgraphics for Windows.

На 1 этапе производится преобразование исходных количественных признаков в серию бинарных (в проводимых нами исследованиях использовался модернизированный алгоритм адаптивного кодирования разнотипной информации).

На 2 этапе вводятся в альтернативные классы "шумящие" объекты, равномерно распределенные в исходном пространстве признаков. Шумящий класс представляет собой множество фальсификаторов, "столкновение" с которыми способствует лучшему проявлению устойчивых логических закономерностей в данных.

На 3 этапе определяется наиболее "перспективный" объект, относительно которого будет происходить центрирование данных (построение локального пространства). Это реализуется с помощью отображения данных на плоскости двух первых главных компонент (метода главных компонент), принадлежащих одному из диагностических клас-

сов. В качестве центрального может выбираться объект, являющийся как "типичным" представителем своего диагностического класса, так и расположенный на границе облака рассеяния.

На 4 этапе проводится центрирование данных относительно выбранного объекта.

На 5 этапе с помощью процедуры множественной регрессии по методу наименьших квадратов рассчитываются параметры новой координатной оси в локальном пространстве, максимально коррелирующей с зависимой переменной.

На 6 этапе из построенной регрессионной модели исключается константа, и удаляются признаки с отрицательными весами.

На 7 этапе из регрессионной модели средствами интерактивной графики исключаются наиболее удаленные от прогнозируемых значений объекты.

На 8 этапе анализируются весовые коэффициенты признаков, входящих в регрессионную модель. Признаки с минимальными весовыми коэффициентами исключаются из дальнейшего анализа.

Седьмой и восьмой этапы повторяются до тех пор, пока весовые коэффициенты признаков, входящих в регрессионную модель не примут значения единицы. Построенная таким образом регрессионная модель может быть представлена в виде логического правила "Если ... То ...". Следующий шаг состоит в определении точности и полноты построенного логического правила на имеющейся обучающей или контрольной выборке. Результаты заносятся в итоговую таблицу.

Далее повторяются этапы с третьего по восьмой, пока не будут рассмотрены все "перспективные" объекты исследуемого диагностического класса.

После того, как будут получены диагностические правила с их оценками эффективности, совместный анализ может выявить, что для отдельных объектов одновременно срабатывают два и более правил, в том числе и из разных диагностических классов. В данном случае предпочтение отдается правилу, обладающему наибольшей полнотой и точностью.

Следует отметить, что данный метод требует несколько больших временных затрат, чем методы, которые будут представлены ниже. Однако, он "гарантированно приводит" к лучшим результатам, по сравнению с классическими методами многомерного анализа данных [7], в случае, когда границы диагностических классов не размыты, исследуемых переменных порядка тридцати и объектов не более 150. В противном случае работа с выделением логических правил становится достаточно трудоемкой.

3.2. Построение логических моделей методом ограниченного перебора

Система WizWhy предприятия WizSoft является современным представителем подхода, реализующего ограниченный перебор (максимальная длина конъюнкции элементарных событий равна 6). Исходные данные могут храниться в различных фор-

матах (ASCII, dBase, MS Access, MS Excel и др.). И после загрузки системы остается лишь выбрать нужный формат данных и загрузить данные. После этого, задаются параметры поиска логических правил.

К ним относятся: точность для "Если ... То" (*If-Then*) и для "Если ... То Не" (*If-Then-Not*) правил; максимальное количество событий в правиле; минимальное число объектов, на которых подтверждается правило и запускается процесс поиска правил.

Результаты поиска разделены на несколько отчетов.

Во-первых, это отчет с обнаруженными логическими правилами, где указывается: общее число обработанных объектов, заданные параметры поиска, априорная вероятность прогнозируемого значения переменной. Кроме того, указывается прогнозирующая способность всей совокупности обнаруженных правил, а именно: количество пропусков при прогнозировании, число ложных тревог, общее количество ошибок, общие потери, вероятность успешного прогнозирования для целевого класса, вероятность прогнозирования альтернативного класса, количество объектов, не охваченных выделенными правилами.

Во-вторых, приводится отчет о трендах, в котором представляются результаты сегментации признаков с указанием информативности каждого сегмента.

В третьих, приводится отчет о неожиданных правилах. Под неожиданными правилами в системе WizWhy понимаются правила в виде конъюнкции двух и более простых высказываний, комбинация которых дает точность и полноту прогноза выше, чем это можно было бы ожидать при независимости простых высказываний.

На основе полученных правил в системе WizWhy предусмотрено два варианта предсказания значений целевого признака на новом материале:

- 1) ввод непосредственно одного объекта,
- 2) обработка массива новой информации.

Среди правил могут быть варианты прогноза для двух альтернативных классов, в данном случае "побеждает" правило, имеющее более высокую точность и полноту.

К недостаткам данной системы можно отнести следующее: процесс поиска логических закономерностей сильно растянут во времени и зависит от мощности используемого компьютера; при объемах выборки более 150 объектов и порядка 100 признаков выдача результатов поиска может занять более часа.

Кроме того, система WizWhy выдает, действительно, заявленное авторами "максимальное" число логических правил. В частности, были получены более 12000 правил, среди которых встречались правила, обладающие низкими значениями точности и полноты. Разработчикам следовало бы внести ограничение на вывод правил, обладающих вероятностью ошибки более 5 %.

3.3. Построение логических моделей с использованием деревьев решений

Деревья решений являются достаточно распространенным подходом к выявлению и изображению логических закономерностей в данных. В данном исследовании для построения деревьев решений использовалась одна из наиболее известных систем See5, реализующей модифицированный алгоритм C4.5. Результаты можно представить в виде деревьев решений или в виде логических правил.

Дерево решений состоит из вершин, ветвей, узлов (мест разветвления), листьев. Самая первая (верхняя) вершина называется корнем. Цепочка "корень — ветвь — вершина — ... — вершина" заканчивается вершиной, которую называют "листом". Из каждой внутренней вершины (т.е. не листа) может выходить две или более ветвей. Каждому узлу сопоставлен некоторый признак, а ветвям — либо конкретные значения для качественных признаков, либо области значений для количественных признаков.

Качество дерева характеризуют два основных показателя: точность и сложность дерева. Точность дерева показывает насколько хорошо разделены объекты разных классов. К показателям сложности дерева относятся число листьев дерева, число его внутренних вершин, максимальная длина пути из корня в конечную вершину и др. Показатели сложности и точности взаимосвязаны: чем сложнее дерево, тем оно, как правило, точнее.

Для того, чтобы начать работу с системой See5, необходимо подготовить данные в нужном формате. Система требует задание двух обязательных файлов: первый с именами переменных и заданием классификационной переменной (*.names) и второй с данными (*.data), где по строкам располагаются объекты, а по столбцам признаки, причем в том порядке, в котором они указаны в файле названий.

После загрузки данных задаются параметры конструктора и запускается процесс построения дерева решений. В качестве параметров могут задаваться: перевод деревьев решений в коллекцию правил, объединение отдельных значений в подмножества для уменьшения ветвления дерева, построения леса решений, задание нечетких границ подмножеств значений, процент допустимой ошибочной классификации и минимальное количество объектов, соответствующих выделенному листу.

Каждое правило содержит следующие параметры: количество объектов, соответствующих данному правилу, количество объектов, не принадлежащих данному диагностическому классу, точность правила и вероятность правильного распознавания. В случае, когда срабатывает несколько правил для одного объекта, выбирается диагностический класс, обладающий наибольшим весом.

Когда строится лес решений, то выводятся все построенные деревья решений и их характеристики. Результатом для леса решений является общая точность классификации.

На основе построенных классификаторов (деревьев решений и логических правил) можно проводить классификацию и прогнозирование для новых объектов. Исследователю предлагается ввести значения признаков, находящихся в вершинах дерева, начиная с корня. Результатом является вероятность правильной классификации к одному из диагностических классов.

В системе также предусмотрена классификация множества объектов. Для этого формируется файл (*.cases), аналогичный файлу с данными. Результат классификации представлен в виде списка объектов с проставленными диагностическими классами. Выбор объекта позволяет просмотреть значения признаков, а также правило, которое для него сработало.

Данная система обладает несомненным достоинством: процедура поиска занимает всего несколько секунд. Но, как правило, получаемые диагностические решения имеют не более пяти-шести конъюнкций элементарных событий.

4. Построение метаструктуры диагностических решений

Построение метаструктуры подразумевает выделение элементов предметной области, определение их взаимосвязей и семантических отношений. Рассматривая сконструированные логические правила как отдельные понятия исследуемой предметной области, следует создать систему (совокупность) понятий. Данная система должна обладать следующими свойствами:

- 1) уникальностью или отсутствием избыточности;
- 2) полнотой описания предметной области;
- 3) валидностью (достоверностью) — соответствием выделенных единиц смысловой информации;
- 4) непротиворечивостью.

Установление взаимосвязей предполагает установление семантической близости между отдельными понятиями. Для того, чтобы установить взаимосвязь между логическими правилами, следует выделить группы правил, в которых отдельные элементы появляются с определенной регулярностью. Внутри выделенных таким образом группировок устанавливаются ассоциативные взаимосвязи.

Последним этапом построения модели предметной области при концептуальном анализе является установление семантических отношений между выделенными понятиями и метапонятиями. Установить семантические отношения — это значит определить специфику взаимосвязи, полученной в результате применения тех или иных методов. Для этого строится дерево, вершинами которого являются концепты, а дугами — связки (например, "приводит к", "обуславливая", "сочетаясь", "определяет" и т.д.). Этот позволяет установить наряду с базовыми от-

ношениями, отношения, специфические для конкретной предметной области.

Еще один способ выявления метапонятий состоит в следующем. Если логические закономерности выявлять в локальном пространстве "типичных" объектов рассматриваемых диагностических классов, то полученные таким образом правила можно рассматривать как расстояния от центральных объектов в их собственных локальных метриках до всех остальных объектов выборки. Затем с помощью того или иного алгоритма иерархической кластеризации определяется, что общего, и что разного у полученных кластеров [16].

Заключение

На наш взгляд, только всестороннее изучение имеющейся проблемы даст возможность находить оптимальное диагностическое решение. Использование представленной технологии конструирования диагностических решений и последующее построение метаструктуры диагностических решений позволяет выявлять скрытые логические закономерности исследуемой проблемной области. Рассмотренные

методы могут дополняться или замещаться другими методами поиска скрытых закономерностей, однако это потребует пересмотра процедуры построения метаструктуры получаемых иным способом диагностических решений. Построение метаструктуры является весьма существенным для построения баз знаний, требующих ввода понятий, метапонятий и семантических отношений на основе множества фрагментов знаний о предметной области.

Несмотря на имеющиеся недостатки рассмотренных методов построения диагностических решений в виде логических правил авторам удалось выявить устойчивые закономерности в данных и тем самым расширить базу знаний при решении задач определения специфики межпроцессуальных взаимосвязей когнитивных процессов и исследования специфики организации ментального опыта субъектов с высоким уровнем интеллектуального развития.

Работа частично поддержана грантами РФФИ, проекты №01-01-00772, №03-01-06115, №03-06-80128 и РГНФ, проект №01-06-00084а.

СПИСОК ЛИТЕРАТУРЫ

1. Загоруйко Н.Г. Прикладные методы анализа данных и знаний. — Новосибирск: Изд-во Ин-та математики, 1999. — 270 с.
2. Гладун В.П. Эвристический поиск в сложных средах. — Киев: Наукова Думка, 1977. — 206 с.
3. Финн В.К. Интеллектуальные системы и общество. Сборник статей. — М.: РГГУ, 2000. — 160 с.
4. Поспелов Д.А. Логико-лингвистические модели в системах управления. — М.: Энергоиздат, 1981. — 232 с.
5. Лбов Г.С., Старцева Н.Г. Логические решающие функции и вопросы статистической устойчивости решений. — Новосибирск: Изд-во Ин-та математики, 1999. — 212 с.
6. Лбов Г.С. Методы обработки разнотипных экспериментальных данных. — Новосибирск: Наука, 1981. — 159 с.
7. Дюк В.А. Обработка данных на ПК в примерах: Статистические расчеты. Построение графиков и диаграмм. Анализ данных. — СПб.: Питер, 1997. — 240 с.
8. Барабаш Б.А. Минимизация описания в задачах автоматического распознавания образов // Техническая кибернетика. — 1964. — № 3. — С. 32–44.
9. Marill T., Green D.M. On the effectiveness of receptors in recognition systems // IEEE Trans. V. IT-9. — 1963. — P. 11–17.
10. Kittler J.A. Feature set search algorithms // Proc. Conf. On Pattern Recogn. and Signal Processing. — Paris, France, 25 June–4 July, 1978. — P. 41–60.
11. Кофман А. Введение в прикладную комбинаторику. — М.: Наука, 1975. — 286 с.
12. Бонгард М.М. Проблема узнавания. — М.: Наука, 1967. — 220 с.
13. Янковская А.Е., Берестнева О.Г., Муратова Е.А. Эффективный алгоритм адаптивного кодирования разнотипной информации // Искусственный интеллект в XXI веке: Труды Междунар. конгр. Том 1. — М.: Физматлит, 2001. — С. 155–166.
14. Таранова Н.Н. Метод адаптивного кодирования признаков // Динамика систем. Межвуз. тематич. сб. научн. тр. / Под ред. Ю.И. Неймарка: Нижегород. гос. ун-т, Нижний Новгород, 1995. — С. 54–70.
15. Янковская А.Е., Алексеева Н.В. Извлечение знаний с учетом целей и проблемной области создаваемой экспертной системы // Теория и применение искусственного интеллекта: Сб. трудов второго международного научного семинара. — Сазопол, НРБ, 1989. — С. 387–392.
16. Дюк В., Самойленко А. Data Mining: Учебный курс. — СПб.: Питер, 2001. — 368 с.