
РОЛЬ АНАЛИЗА КОМПЬЮТЕРНЫХ ДИАГНОСТИЧЕСКИХ ПРАВИЛ В ПОДДЕРЖКЕ ПРОЦЕССА РАСШИРЕНИЯ ВРАЧЕБНЫХ ЗНАНИЙ

Б.А. Саруханов

Кафедра информационных систем
Факультет кибернетики
Московский государственный технический университет
радиотехники электроники и автоматики
пр. Вернадского, 78, Москва, Россия, 117454

Предложен алгоритм кластеризации компьютерных диагностических правил, который обеспечивает возможность интерпретации правил в качестве средства расширения знаний специалиста.

Ключевые слова: врожденный сифилис, расширение знаний специалиста, кластеризация, системы поддержки принятия решений.

Принятие решений во многих областях человеческой деятельности сопровождается необходимостью анализа множества факторов. Подобные ситуации характерны и для медицины. Примером такой медицинской задачи может служить диагностика состояния новорожденных от матерей, перенесших инфекционное заболевание (сифилис) [1]. Отсутствие документального подтверждения лечения матери, невозможность полноценного динамического наблюдения за ребенком, что достаточно часто встречается у женщин, перенесших сифилис [2], — все эти факторы затрудняют диагностику состояния новорожденного. Это, в свою очередь, диктует необходимость расширения ресурса знаний, имеющихся у эксперта, которые помогут эффективно принимать решения в задачах, ранее вызывавших затруднения.

В качестве решения проблемы повышения эффективности врачебной диагностики был предложен принцип «машинного обучения на примерах» методом перебора конъюнкций — один из методов интеллектуального анализа данных (ИАД) [3]. Для формирования решения по такому принципу требуется набор объектов с известными характеристиками и с известным экспертным диагнозом. Такой набор называется материалом обучения, для которого строится проблемно-ориентированная база данных (ПОБД) [4]. На основе примеров из этой базы данных алгоритмы ИАД формируют набор диагностических правил (ДП) для каждого из сравниваемых классов состояний. Он называется компьютерной диагностической базой знаний (КДБЗ). Такая база знаний, в принципе, уже может использоваться в различных пользовательских консультационно-диагностических системах.

Однако эксперту для расширения своих профессиональных знаний важна не только информация о каждом ДП в отдельности, но и сведения об их возможных взаимосвязях. В этом случае эксперт может интерпретировать не только каждое правило, но и содержательно оценивать группу близких (часто встречающихся вместе) ДП.

Несмотря на положительные результаты применения компьютеров для решения различных классов слабо формализованных задач, построенные компьютерные базы знаний не использовались полностью. Это объясняется тем, что компьютерные реализации, в основном, были ориентированы на достижение эффективной диагностики, и перед ними не ставилась задача «расшифровки» полученных диагностических правил на профессиональном языке пользователя.

Целью представленной работы явилось обеспечение поддержки процесса расширения профессиональных знаний врача-эксперта с использованием построенных компьютерных диагностических правил.

Данные о методике исследования. На рис. 1 отображен весь путь решения проблемы, возникшей у специалиста: обсуждение самой проблемы и способов ее решения с аналитиком (когнитологом), формулирование задач, выбор способов решения и представления результатов анализа.

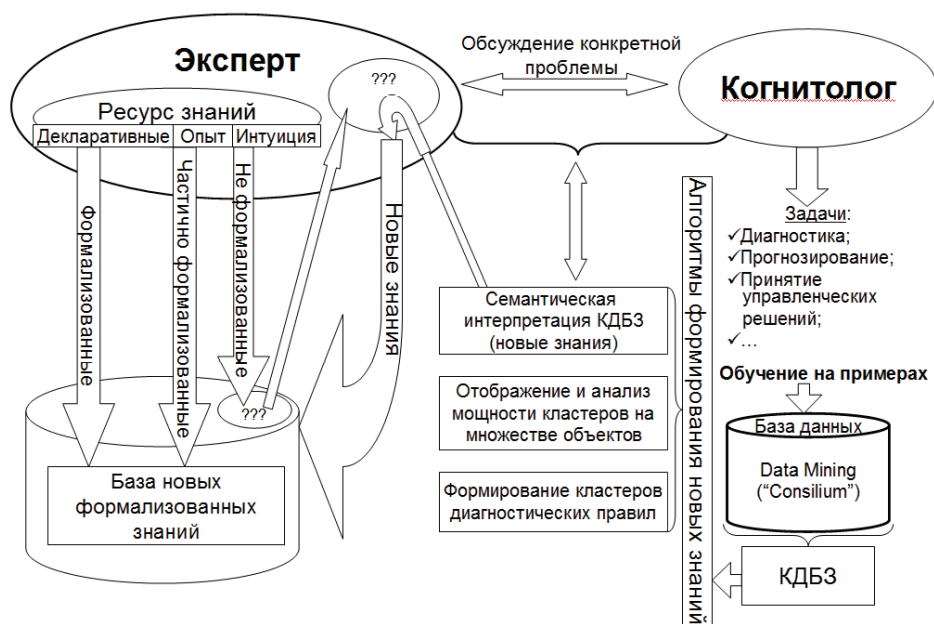


Рис. 1. Концепция расширения знаний эксперта

Для более полного семантического (на профессиональном языке) понимания полученных ДП возникла необходимость разработки методов дополнительного анализа построенных компьютерных баз знаний и алгоритмов, обеспечивающих максимальную интерпретацию совокупности выявленных закономерностей. Такой подход ориентирован на проведение содержательной интерпретации выявленных диагностических правил, обеспечивающих расширение профессиональных знаний врача-эксперта.

Для этого был разработан пакет алгоритмов «Система анализа базы знаний» (САБЗ) для выявления и учета связей между ДП в КДБЗ, группирования их и визуального представления выявленных групп, обеспечивающих семантическую интерпретацию построенных знаний [5; 6].

Алгоритм базируется на идее, что симптом, общий для некоторого количества ДП, может служить основой для группирования таких правил в качестве содержательно близких в данной предметной области. Все симптомы участвуют в исследовании под своими индивидуальными номерами, что позволяет идентифицировать их в протоколах программы с помощью таблицы соответствия.

Применение алгоритма САБЗ позволяет выявлять группы правил (фрагменты КДБЗ), содержательно близких между собой. Данные группы называются кластерами. Построение кластеров ведется до тех пор, пока не будут учтены все имеющиеся в базе знаний правила. Стоит отметить, что ДП, уже попавшие в один из кластеров, не подвергаются повторному анализу на наличие связей. Таким образом, каждое правило попадает только в один кластер. Анализ кластеров обеспечивает получение более полной картины ситуации в данной части проблемной области.

На рис. 2 представлен наиболее информативный (обобщающий) кластер, выявленный алгоритмом САБЗ для одного из классов сравнения («Неблагополучный исход беременности»). Кластер состоит из трех частей (подкластеров), они обозначены овалами и пронумерованы (1, 2, 3).

| | | | | | | |
|--------|-----|------------|------------|------------|-----|--------|
| (10-0) | 197 | 61 | | | | |
| (14-0) | 191 | 99 | | | | |
| (16-0) | 32 | 268 | | 376 | 376 | (14-0) |
| (17-0) | 197 | 77 | | 126 | 77 | (30-0) |
| (18-0) | 77 | 94 | | 126 | 126 | (32-0) |
| | | 55 | | 59 | | |
| | | 53 (44-0) | | 53 (44-0) | | |
| | | 258 (16-0) | 288 (47-0) | 48 (41-0) | | |
| | | | | 243 (10-0) | | |
| | | 77 (41-0) | 245 (41-0) | | | |
| | | | 287 (30-0) | 376 (14-0) | | |
| | | | 237 (59-2) | 4 (13-0) | | |
| | | | 5 (18-0) | | | |
| | | | 27 (77-9) | 305 (11-0) | | |
| | | | 4 (16-0) | | | |
| | | | 305 (13-0) | 37 (12-0) | | |
| | | | 303 (13-0) | | | |
| | | | 371 (12-0) | | | |

Рис. 2. Кластер диагностических правил за класс «Неблагополучный исход беременности»

В ячейках подкластера указаны номера симптомов, входящих в него. В скобках указано количество объектов за оба класса состояний, у которых проявляется ДП. Например, в первом подкластере присутствуют симптомы 55, 59 и 237, комбинация которых встречается у 59 объектов из неблагоприятного класса и у 2 объектов из благополучного.

Подкластер № 2 основан на симптоме под номером 55. Слева над ним расположены номера симптомов, входящих в этот подкластер. Слева от них в скобках

указаны количества объектов, у которых проявляются комбинации этих симптомов. Подкластер № 3 построен симметрично для симптома № 59.

Из рис. 2 следует, что в основе изображенного кластера лежат симптомы № 55 («мать не лечилась до беременности») и № 59 («мать не лечилась во время беременности»). Алгоритм выявил в данных правилах симптом № 53 («мать не лечилась в течение всего периода наблюдения»), что семантически тождественно объединению симптомов 55 и 59. Это позволило провести содержательное преобразование КДБЗ, снизив опасность необоснованного перевеса голосов в сторону неблагоприятного исхода беременности.

Также алгоритмом были выявлены группы правил, которые проявлялись у пациентов, не лечившихся только в один из двух периодов наблюдений (либо до беременности (№55), либо во время (№59)).

Наличие данных групп показало, что форма представления содержательной структуры проблемной области в виде кластеров позволяет выявлять дополнительные связи между симптомами, оказавшимися в разных правилах одного кластера. Это облегчает задачу эксперту, ведущему построение обобщенной картины проблемной области. Анализ и интерпретация содержания всех выявленных кластеров проводится аналогичным образом. Результат интерпретации врачами-экспертами полученных кластеров приведен в табл. 1.

Таблица 1

Пример интерпретации врачом-экспертом кластеров правил

| № кластера | Текст |
|------------|---|
| 1 | У матерей при отсутствии дородового лечения (53) сохраняются высокие титры н-трепонемных тестов ($n/t \geq 5$) в околородовый период (237, 245), что объясняет их высокие значения (≥ 5) у новорожденного (289, 297). Особенно ярко это проявляется у молодых матерей (до 21 года) (4, 5), причем для новорожденных характерно — высокие индексы n/t (≥ 5), с уровнем их значений равным или даже выше материнских (303, 305, 371). |
| 2 | Причиной раннего родоразрешения (31—35 нед.) (94) может служить: отсутствие специфического лечения матери до беременности (55), неадекватно проведенное лечение во время беременности (61), что исключило возможность проведения эффективного профилактического лечения (77). Характерно, что в этих случаях n/t у матерей во втором триместре беременности и после родов имеют резко позитивные значения (191, 197, 268). |
| 3 | Отсутствие у матери специфического и профилактического лечения (77) может привести к рождению детей с результатами индекса значений n/t выше, чем у матери (376), при явно выраженных клинических проявлениях у новорожденного (126). |

Заключение. Разработанный алгоритм кластеризации и выявление причинно-следственных связей диагностических правил в проблеме ВС дают специалисту возможность по-новому взглянуть на знания предметной области. Это, в свою очередь, обеспечивает не только правильную постановку диагноза, но и ее аргументацию, снижая риск врачебной ошибки.

Практическое использование всей совокупности кластеров предоставляет специалисту возможность формирования обобщенной картины проблемной области.

Алгоритм кластеризации диагностических правил может рассматриваться как одно из средств поддержки процесса расширения профессиональных знаний специалиста.

Применение данного алгоритма возможно и в других областях человеческой деятельности.

ЛИТЕРАТУРА

- [1] *Шувалова Т.М., Борисенко К.К.* К вопросу о клинике и диагностике раннего врожденного сифилиса // ИППП. — 1999. — № 4. — С. 13—18.
- [2] *Тихонова Л.И., Сафарова Э.А., Салахов Э.Р.* Масштаб распространения, определяющие факторы и последствия врожденного сифилиса в Российской Федерации // Вестник дерматологии и венерологии. — 2006. — № 6. — С. 41—43.
- [3] *Карп В.П.* Интеллектуальный анализ данных в проблеме построения решающих правил классификации (на примере задач медицинской диагностики) // Новости искусственного интеллекта. — 2006. — № 2. — С. 57—75.
- [4] *Карп В.П., Саруханов Б.А.* Формирование проблемно-ориентированной базы данных в условиях пассивного эксперимента // Сборник трудов научно-технической конференции. Ч. 1. «Информационные технологии и системы. Вычислительная техника». МИРЭА. — М., 2007. — С. 53—57.
- [5] *Карп В.П.* Интеллектуальный анализ данных в проблеме формализации интуитивных знаний эксперта. Сб. трудов конференции «Нелинейная динамика в когнитивных исследованиях» — 2011: труды конференции / Рос. акад. Наук, Ин-т приклад. физики. — Нижний Новгород: ИПФ РАН, 2011. — С. 82—85.
- [6] *Карп В.П., Саруханов Б.А.* «Система анализа базы знаний» (САБЗ), Свидетельство о государственной регистрации программы для ЭВМ № 2011617361, г. Москва, 2011 г.

THE ROLE OF COMPUTER DIAGNOSTIC RULES ANALYSIS IN SUPPORT OF A MEDICAL KNOWLEDGE EXPANSION PROCESS

B.A. Sarookhanov

Information Systems (sub)department

Cybernetics Department

Moscow State Technical University of Radioengineering Electronics and Automation

Vernadsky ave., 78, Moscow, Russia, 117454

An algorithm of computer diagnostic rules clusterization is presented. It supports a possibility of rules interpretation as specialist's knowledge expansion tool.

Key words: congenital syphilis, specialist's knowledge expansion, clusterization, decision support systems.