

ГРАФИЧЕСКАЯ ВИЗУАЛИЗАЦИЯ ДАННЫХ, ЗНАНИЙ И ЗАКОНОМЕРНОСТЕЙ В ПРИКЛАДНЫХ ИНТЕЛЛЕКТУАЛЬНЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

Янковская А.Е., Аметов Р.В., Черногорюк Г.Э.¹

Томский государственный архитектурно-строительный университет,
Лаборатория интеллектуальных систем
634003, Россия, Томск-3, Соляная пл.2
e-mail: yank@tisi.tomsk.su

¹Сибирский государственный медицинский университет, Россия
Работа поддержана Российским Фондом Фундаментальных Исследований, проекты № 98-01-00295, 98-01-03019.

ABSTRACT

In this paper different means for graphical representation and visualisation of data and knowledge in applied intelligent information systems are proposed. Graphical means of two types with and without mapping in usual reality are developed. The results of scientific researching and practical using of this algorithms are given.

ВВЕДЕНИЕ

По мере возрастания интереса к созданию и применению интеллектуальных прикладных информационных систем (ПИИС) повышается интерес к использованию в них графических средств. Графические образы на экране дисплея способствуют более плодотворному общению с пользователем системы, пониманию им информационных структур [1], описывающих данные и знания в ПИИС, и восприятию различного рода выявляемых в процессе решения закономерностей, а также визуализации решения задачи.

Данная работа отражает развитие описанных в [2,3] графических средств представления знаний и закономерностей. В ней обосновывается целесообразность применения в ПИИС графических средств, имеющих и не имеющих отображения в обычной реальности. Развиваются предложенные в публикациях [2,3] не имеющие отображения в обычной (повседневной) реальности графические средства визуализации информационных структур и выявленных закономерностей в данных и знаниях, представленных в матричном виде [3]. Кроме того, предлагается образное отображение данных и знаний о конкретной проблемной области на примере заболевания бронхиальной астмой. В целях понимания изложения графических средств вводятся понятия данных, знаний, закономерностей, приводится используемый способ представления данных и знаний. Изложение сопровождается иллюстративными примерами.

ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ. СПОСОБ ПРЕДСТАВЛЕНИЯ ДАННЫХ И ЗНАНИЙ

В целях изложения графических средств визуализации информационных структур вкратце опишем нетрадиционный способ представления данных и знаний в виде матриц описаний и различий [2,3]. Строки матрицы описаний сопоставляются объектам из обучающей выборки, столбцы - признакам, характеризующим в совокупности каждый объект. Элемент матрицы описаний задает значение данного признака для соответствующего объекта. Признаки, участвующие в построении матрицы описаний, считаются характеристическими и должны быть двоичными или троичными. Строки матрицы различие-

ний сопоставляются одноименным строкам матрицы описаний, столбцы - признакам, определяющим различные механизмы разбиения объектов на классы эквивалентности (механизмы классификации). Элемент матрицы различений, задает принадлежность данного объекта одному из выделенных классов при соответствующем механизме классификации. Для указания факта принадлежности объекта классу используется номер этого класса. Признаки, участвующие в построении матрицы различений, считаются классификационными и должны быть целочисленными. Множество неповторяющихся строк матрицы различений образует множество выделенных образов. Если имеется единственный механизм классификации, матрица различений вырождается в столбец, что соответствует традиционному представлению знаний в задачах распознавания образов.

С содержательной точки зрения столбцы матрицы различений могут быть любого из следующих трех типов. Первый тип характеризуется вложенностью механизмов классификации, когда каждый последующий столбец задает более подробное разбиение объектов на классы эквивалентности. Второй тип может интерпретироваться как совокупность действий, которые необходимо выполнить для данного объекта (в данной ситуации). Третий тип служит для представления независимых механизмов классификации, где столбцы отражают, например, мнения различных экспертов.

Заметим, что настоящая модель позволяет представлять не только данные, но и знания экспертов, поскольку одной строкой матрицы Q можно задавать в интервальной форме подмножество объектов, для которых характерны одни и те же решения, задаваемые строкой матрицы ψ .

Под закономерностями понимаются подмножества признаков с определенными легко интерпретируемыми свойствами, влияющими на различимость объектов из разных образов, устойчиво наблюдаемыми для объектов из обучающей выборки и проявляющимися на других объектах той же природы, а также весовые коэффициенты признаков, характеризующие их индивидуальный вклад в различимость объектов. К упомянутым подмножествам будем относить константные, устойчивые (константные внутри класса), неинформативные (не различающие ни одной пары объектов и не являющиеся константными, или весовой коэффициент признака меньше наперед заданного значения), альтернативные (в смысле включения в диагностический тест), зависимые (в смысле включения подмножеств различимых пар объектов), несущественные (не входящие ни в один безызбыточный тест), обязательные (входящие во все безызбыточные тесты) признаки [4].

Каждый модуль знаний включает в себя одну матрицу описаний и от 1 до 3 матриц различений, причем, присутствие одной из матриц различений 1-го или 3-го типов обязательно. В зависимости от размерностей матриц описаний и различений, числа механизмов классификаций и классов по каждому из механизмов классификации предлагается полное либо фрагментарное графическое представление структуры модуля знаний.

Приведем иллюстративный пример модуля базы знаний, представленный матрицей описаний и матрицей различений 1-го типа (рис. 1).

		1	2	3	4	5
1	0	1	1	-	0	
2	1	0	-	1	0	
3	1	0	1	0	-	
4	-	1	0	1	0	
5	1	-	1	1	0	
6	-	0	-	0	1	

$Q::$

		1	2	3	4	5
1	1	...				
1	2					
2	2					
2	1					
3	1					
1	1					

$\psi::$

Рис. 1. Нетрадиционный матричный способ представления данных и знаний

ГРАФИЧЕСКИЕ СРЕДСТВА ВИЗУАЛИЗАЦИИ ИНФОРМАЦИОННЫХ СТРУКТУР

Опишем 1-ый способ графического представления модуля базы знаний и представим его для вышеприведенного примера на рис. 2.

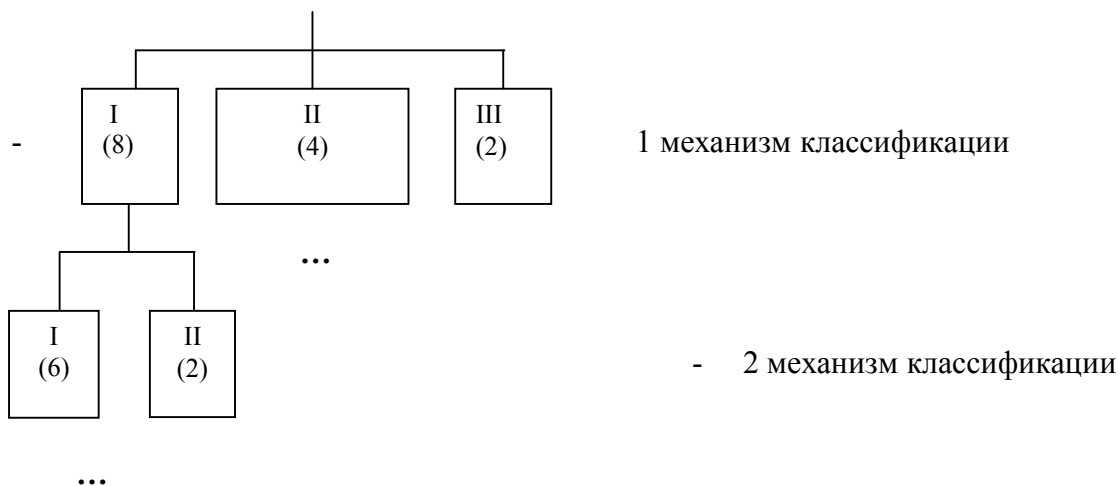


Рис. 2. Пример графического представления данных в модуле базы знаний в виде дерева

Если в модуль базы знаний включена матрица различений 1-го типа, то для ее графического представления используется структура в виде дерева. Вершины i -го уровня сопоставлены i -му механизму классификации, а ребра i -го уровня - классам по i -му механизму классификации. Имя класса с указанием количества объектов, на которых в соответствии с матрицей описаний определен данный класс, отображаются внутри прямоугольника, встроенного в соответствующее ребро дерева. При этом высота прямоугольника существенно больше ширины. Таким образом ребра i -го уровня взвешены количеством объектов соответствующего класса. Для наглядности ширина прямоугольника прямо пропорциональна числу объектов рассматриваемого класса. Простая цепь, соединяющая вершину дерева с листом, определяет решение. Для удобства восприятия классы отображены на рис. 2 римскими цифрами, расшифровка (имя) которых приводится отдельно при необходимости.

Если в модуль знаний включена матрица различений 3-го типа, то для ее графического представления используется структура в виде сети. Для графического представления матриц различения 2-го типа используется бихроматический (двудольный) граф, содержащий в качестве одного подмножества вершин листа дерева либо сети, а другому подмножеству вершин соответствуют различные строки матрицы различений второго типа. Ребро двудольного графа взвешивается числом объектов, на которых в соответствии с матрицей описаний заданы одинаковые строки матрицы различений второго типа. Второй способ графического представления модуля базы знаний отличается от первого расположением на экране дисплея (разворотом на 90°) и в печатном виде графического представления дерева, сети. Все механизмы классификации для матрицы различения 1 (3) типа представляются в порядке следования на одном уровне, причем, расположение прямоугольников 1-го уровня дерева (сети) аналогично 1-му способу, 2-й уровень дерева (сети) располагается справа от 1-го уровня и т.д. Перпендикулярно ребрам проводятся параллельно друг другу линии, число которых равно количеству различных строк матрицы различений 1-го (3-го) типа, т.е. числу различных решений (образов), представленных соответствующей матрицей различения. Ребро соединяется с линией в том случае, если соответствующий ребру класс включен в (сопоставлен) соответствующее линии решение.

Наименование решения с указанием количества объектов обучающей выборки отображается внутри прямоугольника, встроенного в соответствующую линию.

Другой подход к графической визуализации данных и знаний заключается в их представлении с помощью образов, характеризующих их в повседневной реальности. Пример такого представления для исследования больных бронхиальной астмой приведен на рис.3.

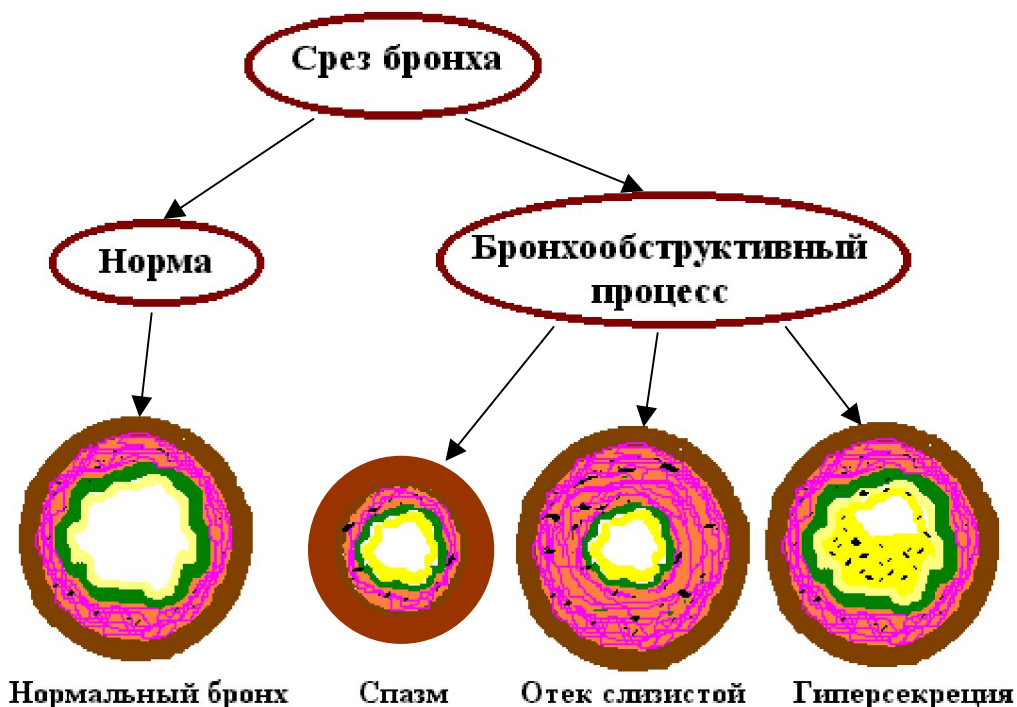


Рис. 3. Отображение данных о бронхообструктивных процессах на примере бронхиальной астмы с использованием древовидной структуры

Специалистами был предложен следующий механизм отображения знаний: существующие в обычной реальности данные и знания представляются в виде деревьев, концевыми вершинами которых служат срезы (сечения) бронха, характеризующие его состояние в целом. Срез закрашивается различными цветами, причем закрашиваемая часть среза прямо пропорциональна величине того или иного бронхообструктивного процесса (спазм гладкой мускулатуры, отек слизистой и гиперсекреция). Образ состояния бронхов складывается у специалиста на основе индивидуального опыта и представления о патогенезе заболевания.

ГРАФИЧЕСКИЕ СРЕДСТВА ВИЗУАЛИЗАЦИИ ЗАКОНОМЕРНОСТЕЙ

Одной из основных подсистем в системе ИИ, основанной на предложенных нами тестовых методах распознавания образов [5,6], является подсистема оптимизации знаний [4,7], предназначенная для анализа модуля базы знаний на репрезентативность, непротиворечивость; выделения константных, неинформативных, альтернативных (в плане одинаковой различимости объектов), зависимых (в плане включения при различении объектов), обязательных (принадлежащих любому тесту) признаков; вычисления весовых коэффициентов признаков; нахождения множества минимальных тестов (подмножеств признаков), обеспечивающих различимость любой пары объектов из обучающей выборки, принадлежащих разным образам. В силу ограниченности рамок доклада приведем основные графические средства визуализации вышеупомянутых закономерностей. Для визуализации соотношений между различными группами признаков (неинформативные, константные, альтернативные, зависимые, обязательные) с указанием номеров признаков,

входящих в соответствующие группы, целесообразно использовать круговую диаграмму. Каждой группе признаков сопоставляется сектор круга, центральный угол которого прямо пропорционален количеству признаков соответствующей группы. Каждый сектор окрашивается заранее выбранным цветом, например, сектор неинформативных признаков - в черный цвет, сектор обязательных признаков - в красный цвет, сектор альтернативных признаков - в голубой, а зависимых - в желтый. Пример визуализации закономерностей с помощью круговой диаграммы приведен на рис. 4.

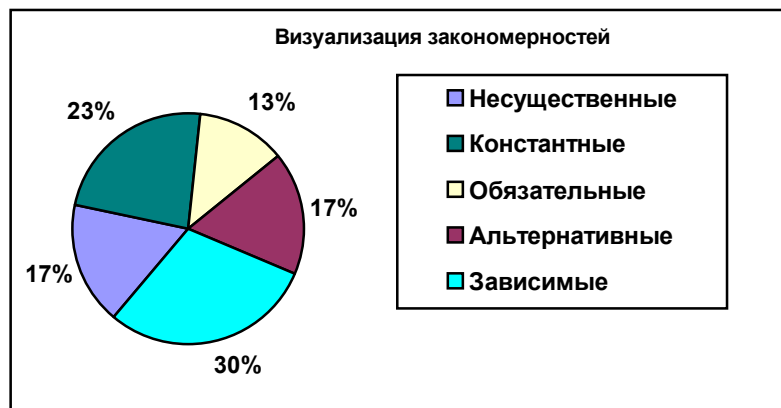


Рис. 4. Визуализация закономерностей с помощью круговой диаграммы

Кроме того, альтернативные признаки представляются двудольным графом, а зависимые - ориентированным графом. Вершины графов отмечаются номерами признаков. Имена признаков по желанию пользователя могут быть представлены на экране дисплея. Множество минимальных подмножеств признаков (минимальных тестов) представляется наглядно с помощью диаграмм Венна. При этом круги, соответствующие разным тестам, окрашиваются в разные цвета и имена признаков того или иного теста, а также обязательные признаки, соответствующие пересечению всех кругов Венна, могут по желанию пользователя выводиться на экран. Заметим, если мощность множества минимальных тестов велика, то представление с помощью диаграмм Венна ненаглядно. Для представления весовых коэффициентов признаков целесообразно воспользоваться гистограммой, на которой высоты прямоугольников прямо пропорциональны весовым коэффициентам сопоставленных им признаков. Эта же информация может быть представлена на разбитой на секторы круговой диаграмме. В каждом секторе вычерчивается дуга с радиусом прямо пропорциональным весовому коэффициенту соответствующего признака. Аналогично вышеописанному имя признака и его весовой коэффициент могут быть выведены на экран дисплея. Для графического представления теста целесообразно использовать дерево (в общем случае сеть) теста, построение которого приводится в [8].

ЗАКЛЮЧЕНИЕ

Предлагаемые графические средства для систем ИИ, основанных на тестовых методах распознавания образов реализованы в системе ИМСЛОГ [9], являющейся инструментальным средством для создания ПИИС и ПИС. Дальнейшее развитие графических средств связано с их ориентацией на пользователей различной квалификации для конкретных проблемных областей и с использованием новых методов принятия решений. Например, таковыми могут быть методы, основанные на применении оптимального сочетания условных и безусловных минимальных диагностических тестов (смешанных тестов).

В заключение отметим, что благодаря использованию наглядных графических средств можно повысить эффективность построенных на тестовых методах распознавания образов обучающих ИС и систем ИИ, предназначенных для практических и научных целей. Образное, имеющее отображение в обычной реальности, представление данных и знаний

позволяет глубже проникнуть в суть процессов, явлений для пользователей прикладных информационных интеллектуальных систем.

ЛИТЕРАТУРА

1. Поспелов Д.А., Литвинов Л.В. Как совместить левое и правое// Новости искусственного интеллекта. -1996.- N2,- с. 66-71.
2. Янковская А.Е. Графические средства в системах искусственного интеллекта, основанных на тестовых методах распознавания образов// Новые информационные технологии в исследовании дискретных структур. Сб. докладов Всероссийской конференции. - Екатеринбург: УрО РАН, 1996. - С. 72-76.
3. Янковская А.Е. Графические средства в интеллектуальных обучающих распознающих системах// Искусственный интеллект в образовании. Труды международного семинара. Часть. 2. - Казань, 1996. - С. 101-106.
4. Янковская А.Е. Тестовые распознающие медицинские экспертные системы с элементами когнитивной графики// Компьютерная хроника. 1994. №№ 8/9. С. 61-83.
5. Янковская А.Е. Оптимизирующие преобразования в процессе синтеза асинхронного автомата и их приложения// MTA SZTAKI TANUL - MANYOK. - BUDAPEST, 99/1980. - С. 212-227.
6. Янковская А.Е. Алгоритмы распознавания на базе оптимизирующих логико-комбинаторных преобразований в пространстве признаков// Математические методы распознавания образов. Тез. докл. IV Всесоюзной конференции. - Рига, 1989. - Часть I. - С. 97-99.
7. Yankovskaya A.Ye., Gedike A.I. Theoretical Base, Realization and Application of the Intelligent System EXAPRAS// Proceeding East-West Conference on Artificial Intelligence "EWAIC'93. From Theory to Practice". - Moscow, 1993. - pp. 248-252.
8. Matrosova A. Yu., Yankovskaya A.Ye. Use of Conditional Nondeterministic Diagnostics Tests in Pattern Recognition Problems// Pattern recognition and Image Analysis. - 1992. - V. 2. -N 2. - pp. 164-168.
9. Гедике А. И., Янковская А.Е. Первая версия междисциплинарной интеллектуальной системы ИМСЛОГ// Труды международной конференции "Интеллектуальное управление: новые интеллектуальные технологии в задачах управления (ICIT'99)", - 1999- С. 32-36