



М. Н. Нессонова

## Метод построения классификаторов на основании геометрической интерпретации структуры данных

*Национальный фармацевтический университет, г. Харьков*

**Ключевые слова:** классификация с обучением, разнотипные признаки, многомерное шкалирование, корреспондентский анализ.

Предлагается метод классификации с обучением, позволяющий построение алгоритмов классификаторов на основании вычисления оценок и базирующийся на метрическом подходе к представлениям взаимосвязей признаков, полученных методами геометрической интерпретации структуры данных.

### Метод побудови класифікаторів на основі геометричної інтерпретації структури даних

*М. М. Нессонова*

Запропонували метод класифікації з навчанням, що дає можливість побудови алгоритмів класифікаторів на основі обчислення оцінок і базується на метричному підході до представлень взаємозв'язків ознак, що отримані методами геометричної інтерпретації структури даних.

**Ключові слова:** класифікація з навчанням, різнотипні ознаки, багатовимірне шкалювання, корреспондентський аналіз.

*Актуальні питання фармацевтичної і медичної науки та практики. – 2014. – № 1 (14). – С. 101–103*

### The method to develop classifiers based on geometrical interpretation of data structure

*M. N. Nessonova*

The method of supervised classification is suggested in the article. The method allows developing mathematical models of algorithms for computing estimates. It is based on metric approach to representations, generated by statistical methods of geometric interpreting of data structure.

**Key words:** supervised classification, heterogeneous features, multidimensional scaling, correspondence analysis.

*Current issues in pharmacy and medicine: science and practice 2014; № 1 (14): 101–103*

Под системами поддержки принятия решений (СППР) в медицине и здравоохранении понимают проблемно-ориентированные системы или программно-аппаратные комплексы, реализующие технологию информационной поддержки процессов принятия лечебно-диагностических и управленческих решений медицинским персоналом [1]. В отличие от информационных систем, автоматизирующих хранение и обработку данных, СППР являются наукоемкими, так как требуют использования методов анализа данных из различных научных направлений. Поскольку величины, содержащиеся в массивах медицинских данных, в большинстве случаев случайны, наиболее адекватными для построения СППР в медицине являются стохастические и вероятностно-статистические модели. Особо актуальны для таких моделей методы многомерного статистического анализа, при помощи которых можно установить характер и структуру взаимосвязей между признаками исследуемого объекта [2]. Одной из наиболее актуальных задач при разработке СППР в медицине на сегодня является формирование решающих правил, согласно которым выполняется процесс диагностики пациентов, а также выделение и формализация информативных структурных элементов, на основании которых синтезируются эти решающие правила [3,4].

Задача классификации с учителем состоит в построении по обучающей информации правил (классификаторов) отнесения новых объектов к одной или нескольким

заданным группам (классам). При этом предполагается, что пространство объектов (генеральная совокупность) разбивается на несколько заранее известных классов. Сведения о пространстве объектов не полны, известно лишь признаковое описание некоторого его подмножества (обучающей выборки). Об объектах обучающей выборки также имеется информация, к какой группе каждый из них принадлежит. Решение задачи классификации с обучением становится основой медицинских СППР, предназначенных для дифференциальной диагностики заболеваний, определения формы заболевания, оценки степени тяжести состояния пациентов, прогнозирования клинического исхода и т.п. Обучающими объектами является выборка пациентов, которые описываются признаками различной природы. В качестве признаков могут выступать значения клинических и лабораторных показателей (количественные признаки), маркеры наличия или отсутствия определенных симптомов (номинальные признаки) или уровня их выраженности (порядковые признаки). Классификатор (решающее правило) при этом представляет собой функцию, зависящую от значений признаков объекта.

#### Цель работы

Разработка метода построения классификаторов, позволяющих использование в качестве предикторов разнотипных признаков с возможным учетом нелинейности и немонотонности их поведения при переходе от класса к классу.



### Материалы и методы исследования

Предлагаемый метод построения классификаторов по обучающей информации основан на метрическом подходе к представлениям (картам), полученным с помощью методов геометрической интерпретации структуры данных. При его разработке использовали методы многомерного шкалирования и корреспондентского анализа. Ключевым моментом в разработке метода стало использование представления о понятии «класс» не как о некоторой компактно расположенной области в пространстве объектов, а как о еще одной компоненте признакового описания. В этой интерпретации строилась математическая модель зависимости целевого признака «класс» от других (объясняющих) переменных.

### Результаты и их обсуждение

При решении практических задач построения классификаторов по обучающей информации рекомендуемый метод классификации предлагается применять в соответствии со схемой (рис. 1).

В представленной на рис. 1 схеме предполагается, что исходная обучающая информация (база данных) подается в виде матрицы признакового описания объектов. На первом этапе происходит структуризация и преобразование таблицы обучения, включающая и предварительную обработку данных (удаление выбросов, зависимых показателей и др.). Исключаются показатели, считающиеся неинформативными для определения принадлежности объектов к классам в силу отсутствия влияния при-

знака «класс» на их значения. Итогом выполнения этих процедур будет получение избыточного набора  $n$  независимых показателей, являющихся потенциальными предикторами классификатора.

Категоризация значений предикторных переменных происходит путем разбиения областей значений количественных показателей на интервалы, характерные для различных классов, а для качественных показателей – объединением тех категорий их значений, которые характерны для одного класса. В результате для каждого класса будет сформирован набор эталонов, представляющих собой элементарные правила сравнения показателей с пороговыми значениями (для количественных признаков), либо перечень категорий, характерных для данного класса (для качественных признаков). При выделении информативных интервалов области значений предиктора исходили из естественных соображений, что переменные, наиболее полно характеризующие различие между классами, по своей природе таковы, что можно разделить диапазоны их изменения таким образом, чтобы частоты встречаемости объектов обучающих выборок этих классов в выделенных интервалах значительно различались.

Итогом первого этапа является обучающая матрица «объект–признак», содержащая  $n+1$  столбец, отвечающий классам объектов и категоризированным значениям  $n$  независимых предикторов. Данная матрица представляет собой входную информацию для получения

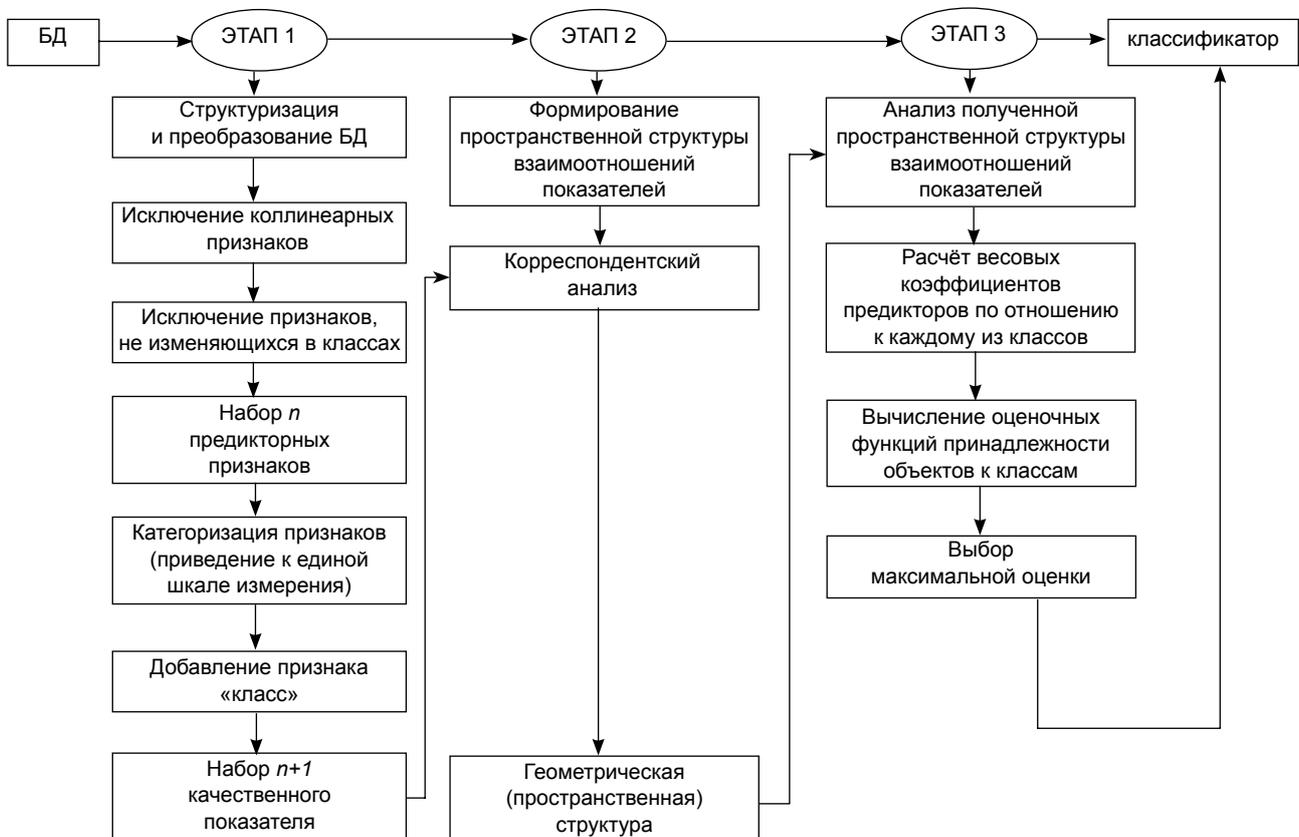


Рис. 1. Алгоритм практического применения разработанного метода классификации.



геометрической интерпретации взаимосвязей между показателями с помощью методов корреспондентского анализа [5] (этап 2).

Если обучающая информация задана в виде таблицы попарных сравнений, то этап предварительной обработки данных, включая приведение переменных к единой шкале, может быть опущен. В этом случае на этапе формирования пространственной структуры взаимосвязей признаков наиболее целесообразно использовать методы многомерного шкалирования [6].

На третьем этапе анализируют полученную пространственную структуру и выбирают метрику, с помощью которой оценивают расстояния между признаками на полученной карте. Мера влияния определенного эталона на принадлежность объекта к конкретному классу оценивается как величина, обратная расстоянию точки-представителя этой категории до точки-представителя этого класса, нормированная на сумму расстояний всех точек-представителей категорий признаков. Для каждого класса вычисляется значение оценочной функции, характеризующей степень принадлежности классифицируемого образца к этому классу. Формулы для вычисления оценочных функций представляют собой взвешенные суммы по всем используемым элементарным логическим правилам. Классифицируемый объект следует относить к тому классу, для которого получено наибольшее значение оценочной функции. Таким образом, формируется общая модель конкретного классификатора.

## Выводы

Применение разработанного метода классификации позволяет повысить качество классификации в медицинских СППР. Математические модели, получаемые на его базе, за счет использования в их основе логических правил обладают достаточно важным свойством интерпретируемости, т.е. позволяют видеть не только окончательный ответ, но и объяснение, почему он был выдан.

## Список литературы

1. Кобринский Б.А. Проблема взаимопонимания: термины и определения в медицинской информатике / Б.А. Кобринский // Врач и информационные технологии. – 2009. – № 1. – С. 51–52.
2. Халафян А.А. Анализ и синтез медицинских систем поддержки принятия решений на основе технологий статистического моделирования : автореф. дис. на соискание ученой степени д.т.н. / А.А. Халафян. – Краснодар, 2010. – 47 с.
3. Ахутин В.М. Оценка качества формализованных медицинских документов / В.М. Ахутин, В.В. Шаповалов, М.О. Иоффе // Медицинская техника. – М., 2001. – №2. – С. 27–31.
4. Поворознюк А.И. Система поддержки принятия решения в медицине на основе синтеза структурированных моделей объектов диагностики / А.И. Поворознюк // Научные ведомости БелГУ. Серия: История. Политология. Экономика. Информатика. – 2009. – №12-1. – Режим доступа: <http://cyberleninka.ru/article/n/sistema-podderzhki-prinyatiyaresheniya-v-meditsine-na-osnove-sinteza-strukturirovannyh-modeley-obektov-dagnostiki>.
5. Greenacre M. Correspondence Analysis in Practice, Second Edition / M. Greenacre. – London : Chapman & Hall/CRC, 2007.
6. Дэйвисон М. Многомерное шкалирование: методы наглядного представления данных / М. Дэйвисон. – М. : Финансы и статистика, 1988. – 254 с.

### Сведения об авторе:

Нессонова М.Н., ассистент каф. фармакоинформатики, Национальный фармацевтический университет,  
E-mail: [saddy\\_me@mail.ru](mailto:saddy_me@mail.ru).

Надійшла в редакцію 04.02.2014 р.

УДК: 378.147:[378.091.5:616.9-053.2]

Т. М. Пахольчук, О. В. Усачова

## Деякі елементи когнітивної технології навчання у викладанні дитячих інфекційних хвороб на V курсі

Запорізький державний медичний університет

**Ключові слова:** когнітивна технологія навчання, викладання дитячих інфекцій.

Описано елементи когнітивної технології навчання у викладанні клінічної дисципліни у вищій школі.

**Некоторые элементы когнитивной технологии обучения в преподавании детских инфекционных болезней на V курсе**

Т. Н. Пахольчук, Е. В. Усачева

Обсуждаются элементы когнитивной технологии обучения в преподавании клинической дисциплины в высшем учебном заведении.

**Ключевые слова:** когнитивная технология обучения, преподавание детских инфекций.

**Актуальные вопросы фармацевтической и медицинской науки и практики.** – 2014. – № 1 (14). – С. 103–105

**Some elements of cognitive technology in the teaching of children's infectious diseases on the V course**

T. N. Paholchuk, H. V. Usachova

The elements of cognitive technologies in the teaching of clinical disciplines in higher school are discussed in the work.

**Key words:** cognitive technology of education, the teaching of children's infections.

**Current issues in pharmacy and medicine: science and practice 2014; № 1 (14): 103–105**