

**А.А. Татаренков**

*Курский государственный технический университет, Россия*

**Е.Н. Корневская**

*Донецкий государственный медицинский университет, Украина*

## **ОБОСНОВАНИЕ ИСПОЛЬЗОВАНИЯ ДИАЛОГОВЫХ СИСТЕМ РАСПОЗНАВАНИЯ ОБРАЗОВ В ЗАДАЧАХ МЕДИЦИНСКОЙ ДИАГНОСТИКИ**

Решение задач диагностики нормальных и патологических состояний человека и окружающей среды может осуществляться известными методами теории распознавания образов в ее геометрической интерпретации, когда объекты, характеризующиеся различными наборами признаков  $X=(X_1, X_2, \dots, X_n)$ , представляются как точки (вектора) многомерного пространства признаков  $N$ . Разделение классов осуществляется с помощью полученных на этапе обучения разделяющих гиперповерхностей.

В настоящее время в практических приложениях, в том числе и в медицинских, используется достаточно большое количество различных методов распознавания образов, причем известно, что большинство из них «хорошо» работают, если структура классификационных данных удовлетворяет ограничениям, присущим соответствующим математическим моделям. Здесь под структурой данных (классов) понимается конфигурация точек-объектов в пространстве признаков, подчеркивая тем самым, что через нее отображаются основные взаимоотношения между исследуемыми объектами. Чтобы решить вопрос о выборе того или иного известного метода распознавания образов, для решения конкретной диагностической задачи (класса задач) требуются хотя бы предварительные сведения о структуре классов решаемой задачи.

Анализ задач медицинской диагностики показал, что многие из них имеют сложную геометрическую структуру. Это связано с тем, что человек и окружающая его среда представляют собой чрезвычайно сложный, динамичный объект, и, следовательно, их состояния могут быть описаны лишь большим набором взаимозависимых признаков, которые с трудом поддаются формализованному описанию.

Кроме того, решение задач автоматизированной диагностики состояния человека и окружающей среды осложняется: разнообразием решаемых задач, имеющих различную структуру классов и признакового описания; отсутствием полного статистического материала, одновременно охватывающего различные стороны функционирования исследуемого объекта; наличием казуистических ситуаций; присутствием в обучающем материале объектов с неточными или ошибочными описаниями; часто «учитель» не имеет возможности точно указать класс объекта на обучающей выборке и не дает информации о наличии переходных зон между классами, хотя в обучающей выборке эти объекты имеются; объекты одного и того же класса в исходном пространстве могут быть представлены различными группами, которые, в свою очередь, распределены среди групп других классов, и т. д.

В этих условиях для успешного решения задачи принятия диагностических решений необходимо иметь аппарат, обеспечивающий изучение структуры классов с выдвижением соответствующих гипотез на языке специалистов исследуемой предметной области. В анализе данных таким аппаратом является вычислительный эксперимент. Для решения задач классификации и диагностики вычислительный эксперимент реализуется с помощью диалоговых интерактивных систем распознавания образов (ДСР). Понятие диалога в этих системах несколько отличается от традиционного. Здесь режим диалога ориентирован, прежде всего, на поддержку пользователем принятия решений о структуре многомерных классов с подбором соответствующих методов и алгоритмов обработки, наиболее подходящих к искомой структуре данных (этап качественного решения задачи анализа данных). На этапе количественного описания данных производится поиск параметров выбранных пользователем моделей, и методом проб и оценок делается окончательный выбор конкретной (чаще всего одной) решающей модели.

Одной из самых сложных и плохо формализуемых задач является задача изучения структуры классов, которая решается путем отображения многомерных данных в одно-, двух- или трехмерные пространства, в которых человек имеет возможность увидеть структурные особенности исследуемых классов, сделать определенные предположения, выдвинуть гипотезы, определить дальнейшую стратегию распознавания и т.д.

При решении задачи отображения многомерных классов в двумерные решающие пространства следует иметь в виду, что всякое снижение размерности ведет к потере информации, однако при правильном выборе критериев и правил отображения можно достичь хороших, с точки зрения пользователей, результатов. На практике широкое распространение получили три основных метода снижения размерности:

1) проектирование данных в различные подпространства, мало искажающие структуру данных, при этом новые оси-признаки выбираются в том же пространстве, что и исходное;

2) конструирование нового пространства описания, оси координат в котором не связаны с исходными признаками (различные типы разверток), причем развертка строится так, чтобы близкие в многомерном пространстве объекты оказались близкими и на разворачивающейся кривой;

3) многомерное шкалирование, предназначенное для поиска таких конфигураций точек в пространстве наблюдений, которые сохраняли бы расстояния в исходном пространстве.

В работе [1] показано, что каждый из этих методов обладает определенными достоинствами и недостатками. Например, для проективных методов справедливо такое ограничивающее правило: далекие на проекции объекты обязательно далеки и в пространстве описания, а близкие объекты в наблюдаемом пространстве на самом деле могут быть далеки в многомерном пространстве. При использовании различных типов разверток, наоборот, близкие на развертке объекты обязательно близки и в исходном пространстве, а далекие на развертке объекты могут быть близки в исходном пространстве. В методах шкалирования хорошие отображающие правила синтезировать достаточно сложно и хороших результатов удается добиться при ограниченном числе исходных признаков (до 20). Поэтому на практике, как правило, либо комбинируют различные методы отображений, либо конструируют специальные способы отображения, которые обеспечивают наглядное представление о свойствах классов, интересующих конкретного пользователя [2].

С точки зрения пользователя, алгоритм работы современных диалоговых систем распознавания образов состоит из следующих основных этапов:

1. С помощью набора различных методов отображения человек, ведущий диалог с ЭВМ, изучает особенности структуры исходных классов на обучающей выборке.
2. Зная ограничения и свойства известных методов распознавания, подбирается набор алгоритмов, который более или менее приспособлен к решению задач с изучаемой структурой данных.
3. Производится расчет параметров для выбранных моделей классификации данных, и методом проб и оценок осуществляется окончательный отбор решающих правил, обеспечивающих решение поставленной задачи с заданным качеством.
4. Полученное правило используется для решения задач диагностики неизвестных объектов в том же пространстве признаков, что и на обучающей выборке.

Реализация такого алгоритма в ДСР приводит к значительному усложнению как программных средств, так и механизмов ведения диалога для человека-оператора, поскольку необходимо обеспечивать хранение и работу различных и достаточно мощных алгоритмов и программ, реализующих процессы обучения, отображения и осуществления диалога. Поэтому эти системы не всегда эффективны, до сих пор малодоступны для широкого круга пользователей, имеют сложное математическое обеспечение, требуют большого статистического обучающего материала и специальной математической подготовки «учителя» и пользователя, что значительно затрудняет диалог.

Для устранения перечисленных и ряда других недостатков предлагается специально разработанный метод динамического интерактивного конструирования двумерных классификационных пространств, который способен сохранять работоспособность при плохо формализованной структуре классов с нарушениями гипотезы о компактности, при большом количестве артефактов в обучающей выборке, в условиях не полностью определенных «учителем» классов, возможной «вложенности» их друг в друга и др. [2].

### **Литература:**

1. Александров, В .В. Анализ данных на ЭВМ (на примере системы СИТО) [Текст] / В .В. Александров, А.И. Алексеев, Н.Д. Горский. М.: Финансы и статистика, 1999.231 с.

2. Корневский, Н.А. Проектирование систем поддержки принятия решений для медико-экологических приложений [Текст]: монография / Н.А. Корневский, В.С. Титов, И.Е Чернецкая; Курск, гос. техн. ун-т. Курск, 2004. 180с.



