

3) при большой глубине дерева объектов входная цепочка разбивается терминальными символами высших уровней (абзацное деление и т.д.).

Рассмотрим пример грамматики объекта «формулы алгебры логики».

Правила (P)	Символ замещения во входной цепочке после применения правила (I)	Функция (F)
1. $\Phi^A \Leftarrow A$	-	-
2. $\Phi^{-A} \Leftarrow \neg A$	A	NOT(A)
3. $\Phi^K \Leftarrow A \wedge A$	A	K(A,A)
4. $\Phi^D \Leftarrow A \vee A$	A	D(A,A)
5. $\Phi^P \Leftarrow (A)$	A	-

$$V_T = \{\neg, \wedge, \vee, ()\} \quad V_N = \{A\} \quad I = \{A\}$$

Правила для объекта «переменная»:

$P^{имя} \Leftarrow a$	A	
$P^{имя} \Leftarrow Aa$	A Присв(имя)	Присв(имя)
$P^{конс} \Leftarrow 0$	C	Присв(знач)
$P^{конс} \Leftarrow 1$	C	Присв(знач)

$$V_T = \{\text{буква или цифра}\} \quad V_N = \{\} \quad I = \{A, C\}$$

**ОПТИМИЗАЦИЯ КОЛИЧЕСТВА ПРИЗНАКОВ И МИНИМИЗАЦИЯ ОПИСАНИЯ РЕШЕНИЯ В ЗАДАЧАХ КЛАССИФИКАЦИИ (РАСПОЗНАВАНИЯ ОБРАЗОВ)**

Тетерин А.Н.

*Удмуртский государственный университет  
Ижевск, Россия*

Распознавание подразумевает проработку широкого комплекса задач, начиная от оцифровки аналоговой информации и заканчивая ее семантической интерпретацией. В этот комплекс обязательно входит последовательное, параллельное или иерархическое решение классификационных задач.

Решение задач классификации имеет самостоятельное значение для корпоративных систем управления, экспертных систем в медицине и экономике при распознавании различных ситуаций, когда по набору заданных признаков (факторов) выявляется сущность некоторой ситуации, в зависимости от которой выбирается определенная последовательность действий. Для этих задач характерны следующие предметные области.

• Интерпретация данных - выбор решения из фиксированного множества альтернатив на базе введенной информации о текущей ситуации.

• Контроль – отклонение в данных о текущей ситуации от плановых целей и нормативов.

• Диагностика - выявление причин, приведших к возникновению ситуации.

• Коррекция - диагностика, дополненная возможностью оценки и рекомендаций действий по исправлению отклонений от нормального состояния рассматриваемых ситуаций.

• Проектирование - определение конфигурации объектов с точки зрения достижения заданных критериев эффективности и ограничений.

• Прогнозирование - предсказание последствий развития текущих ситуаций.

• Мониторинг - контроль с возможной последующей коррекцией. Для этого выполняется диагностика, прогнозирование.

• Управление - мониторинг, дополненный реализацией действий в автоматических системах.

Предлагается три типа верифицированных алгоритмов обучения на ограниченных бесконечных множествах. Первые два основаны на обучении с учителем третий может стать основой новой теории кластерного анализа без поиска центра кластера.

$$R^n - \text{действительное } n\text{-мерное пространство элементов } x = (x_1, x_2, \dots, x_n),$$

$$y = (y_1, y_2, \dots, y_n);$$

$\delta_0 = \inf^{\text{def}} \{ \|x - y\| : x \in A, y \in B \}$  - минимальное расстояние между двумя множествами  $A$  и  $B$ ;

Исключение избыточного признака целесообразно, если  $\delta$  в новом пространстве связано с  $\delta_0$  в предыдущем следующим соотношением:

$$\delta > \sqrt{\varepsilon(2-\varepsilon)}(1-\varepsilon)\delta_0 \text{ для алгоритмов первого типа}$$

$$\delta > \frac{n-1}{n} \sqrt{\frac{n-1}{n}} \delta_0 \text{ для алгоритмов второго типа}$$

$$\delta > \sqrt{\frac{n-2}{n-1}} \left( \frac{\delta_0}{\sqrt{n-1}} \right)^{\frac{1}{n-2}} \sqrt{\varepsilon} \delta_0 \text{ для алгоритмов третьего типа}$$

Аналогично, добавление нового признака оправдано, если:

$$\delta > \left( \sqrt{\varepsilon(2-\varepsilon)}(1-\varepsilon) \right)^{-1} \delta_0 \text{ для алгоритмов первого типа}$$

$$\delta > \frac{n+1}{n} \sqrt{\frac{n+1}{n}} \delta_0 \text{ для алгоритмов второго типа}$$

$$\delta > \sqrt{\frac{n}{n-1}} \left( \frac{\sqrt{n-1}}{\delta_0} \right)^{\frac{1}{n}} \frac{\delta_0}{\sqrt{\varepsilon}} \text{ для алгоритмов третьего типа}$$

Главный недостаток алгоритмов первого типа - результат классификации получается на последних шагах работы алгоритма. Мы его потеряем, если есть ограничения по памяти и времени работы.

Недостаток алгоритмов второго типа - равномерная сходимость. Чем больше шагов, тем меньше они отличаются друг от друга. Отработав 50% времени, мы получаем примерно 50% объема, не обработанного единичного гиперкуба. Достоинство - полиномиальная оценка объема памяти.

Достоинства алгоритмов третьего типа - хорошая сходимость, естественность (не проанализированная область гиперкуба находится вблизи разделяющей два множества границы, распознавание может идти параллельно обучению в фоновом режиме. Отработав 20% времени в двухмерном пространстве, остальные 80% алгоритм затратит на удвоение  $\delta_0$ . Недостаток - экспоненциальная оценка объема памяти. Поэтому необходимо использовать понятие чувствительности и применять алгоритмы первого и второго типа к непроанализированной части гиперкуба.

$$\text{Для алгоритмов второго типа время распознавания } t \approx O\left(\frac{n\sqrt{n}}{\delta_0}\right)$$

$$\text{Для алгоритмов первого типа время распознавания } t \approx O\left(n \log_2 \frac{1}{\delta_0}\right)$$

$$\text{Для алгоритмов третьего типа время распознавания } t \approx O\left(n \log_2 \frac{n-1}{\delta_0}\right)$$

Возникает вопрос, нужен ли  $\log_2 \frac{1}{\delta_0}$  ?

В многомерных пространствах (100,1000...) чаще будет встречаться ситуация  $\delta_0 \approx 1, \delta_0 > 1$ , и это нужно учитывать при проектировании алгоритмов.

Понятие проекции в алгоритмах первого и второго типа позволяет решать задачу минимизации числа признаков (сокращение неинформа-

тивных признаков, не изменяющих количество ячеек) без вычислений по формулам.

Алгоритмы второго и третьего типа могут быть модифицированы для решения задачи дообучения без пересмотра всего обучающего множества. В этом случае обучение и распознавание сливаются в один процесс, качество которого повышается с течением времени, а распознавание начинается с одного элемента обучающего мно-

жества. Для алгоритмов первого типа в этом случае задача обучения решается заново.

Полученные оценки для большого описания множеств оправдывают следующий порядок использования алгоритмов. В качестве первого необходимо использовать алгоритм второго типа, его результаты являются входными данными для алгоритма первого и третьего типа. Их главное отличие друг от друга: первый тип строит разделяющую границу между двумя множествами (в некоторых случаях этого вполне достаточно), третий тип - оболочки разделяемых множеств. Незначительное увеличение времени классификации является небольшой платой за дальнейшую работу с *множествами как самостоятельными объектами с минимальным описанием*. Для получения такого описания потребуются результаты работы алгоритма первого типа.

Общее достоинство теории: для каждой ячейки может быть индивидуально выбран алгоритм и принято решение об изменении пространства признаков. Что в целом позволяет говорить не только о динамическом изменении пространства признаков, но и о динамическом изменении применяемых алгоритмов. Общим критерием изменения пространства признаков (алгоритмы первого и второго типа) можно считать излишнее *дробление*  $R^1$  или, другими словами, количество дочерних листьев (по  $R^1$ ) значительно превосходит количество классов.

#### СПИСОК ЛИТЕРАТУРЫ

1. Тетерин А.Н. Геометрический подход к классификации – новая модель работы нейрона. //ЖВМ и МФ. 1992. Т 31. № 12. С. 1972-1980.

#### *Компьютерное моделирование в науке и технике*

#### **АСПЕКТЫ КОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ В ОЦЕНКЕ ГИПОЛИПИДЕМИЧЕСКОЙ ЭФФЕКТИВНОСТИ У БОЛЬНЫХ ИБС С ПОМОЩЬЮ НЕЙРОСЕТЕВОГО КЛАССИФИКАТОРА NEURO PRO 0.25**

Маль Г.С., Алыменко М.А., Николаев Р.Л.  
ГОУ ВПО «Курский государственный  
медицинский университет Росздрава»  
Курск, Россия

В последние несколько лет наблюдается взрыв интереса к возможности прогнозирования лечебного эффекта с помощью нейронных сетей.

Целью исследования явилась выработка подхода к созданию нейросетевых классификаторов для прогнозирования гипохолестеринемического эффекта эндурацина и выявления значимости факторов, способных повлиять на результат фармакотерапии ИБС.

Методы исследования включали в себя клинические, биохимические, функциональные, статистические (программа Neuro Pro 0.25).

В условиях фармакотерапии эндурацином у больных ИБС с изолированной гиперхолестеринемией можно прогнозировать гипохолестеринемический эффект не менее 14% ( $p < 0,05$ ) у 19% больных, а снижение уровня холестерина до 30% можно прогнозировать не менее, чем у 15% больных и только недостоверный гипохолестерине-

мический эффект около 8% составил не более, чем у 7% больных. Но наряду с гипохолестеринемическим эффектом эндурацин способен оказать гипотриглицеридемический эффект, который имеет аналогичную сопоставимую степень проявления у 12% больных и может приводить к снижению уровня триглицеридов не менее, чем на 19% ( $p < 0,05$ ) у 30% больных.

При лечении эндурацином у больных ИБС с сочетанной гиперхолестеринемией можно прогнозировать гипохолестеринемический эффект не менее 10-12% ( $p < 0,05$ ) у 15% больных, а более 17% ( $p < 0,05$ ) у 24,7% больных и только недостоверный гипохолестеринемический эффект 7% возможен не более, чем у 8% больных. Но наряду с гипохолестеринемическим эффектом эндурацин способен оказать гипотриглицеридемический, который имеет аналогичную степень вероятности проявления, а у 19% больных может приводить к снижению уровня ТГ на 25%.

Таким образом, проведенное исследование показало возможность прогнозирования степени гипохолестеринемического эффекта у больных ИБС с изолированной или сочетанной гиперхолестеринемией в условиях фармакокоррекции эндурацином, что необходимо для практической работы врача, так как начиная лечение, необходимо быть уверенным в возможности достижения клинического эффекта.