

7. *Greenland, S. and Poole, C. Invariants and noninvariants in the concept of interdependent effects // Scand. J. Work Environ. Health. – 1988. – Vol.14. – P.125–129.*

E-mail: vpanov@ecko.uran.ru.

УДК 576.89:616.9

Г.Ш. Цициашвили, д-р физ.-мат. наук, **Е.И. Болотин**, д-р биол. наук
(Институт прикладной математики ДВО РАН,
Тихоокеанский институт географии ДВО РАН, Владивосток)

РАЗРАБОТКА БЫСТРОГО АЛГОРИТМА РАСПОЗНАВАНИЯ ОБРАЗОВ В ПРИЛОЖЕНИИ К ЗАДАЧАМ ПРОГНОЗИРОВАНИЯ

Представляется быстрый алгоритм распознавания объектов, у которых основной признак превышает некоторый критический уровень. Этот алгоритм основан на прямоугольном решающем правиле в пространстве значений факторов. Рассматривается возможность адаптации данного решающего правила к сфере факторного прогнозирования.

Ключевые слова: распознавание объектов, критический уровень, факторное прогнозирование.

В настоящей работе представляется быстрый алгоритм распознавания объектов, у которых основной признак (например, годовая заболеваемость той или иной инфекционной патологией) превышает некоторый критический уровень или эпидемический порог. Распознавание ведется по совокупности факторов, характеризующих распознаваемый объект (например, метеорологические условия года). Предлагаемый алгоритм основан на прямоугольном решающем правиле в пространстве значений факторов. Этот алгоритм возник после длительного анализа различных вариантов решения задачи распознавания образов применительно к выявлению критических значений основного признака.

В распознавании образов выделяются два основных подхода. Первый из них основывается на линейной разделяющей плоскости, строящейся в пространстве факторов. Этот подход требует для своей реализации нахождения оптимальной разделяющей плоскости, что само по себе делает задачу весьма непростой в вычислительном плане и требует достаточно большого числа арифметических операций. Второй подход основан на методе тупиковых тестов и применяется в том случае, когда факторы принимают значение 0, 1. В этом случае выделение всех тупиковых тестов становится, строго говоря, NP-полной задачей перебора. Поэтому данный вариант очень непрост в компьютерной реализации. Если же отказаться от метода распознавания образов и обратиться к процедурам многомерной статистики, то в приложениях возникает ситуация, называемая «проблемой малых выборок» когда число наблюдаемых объектов относительно невелико (порядка 10-30), а число факторов достаточно велико (5 и более). В то же время во всех пособиях по математической статистике рассматривается противоположная ситуация, когда число факторов невелико, а число объектов – достаточно велико.

Основная идея построения прямоугольного решающего правила состоит в определении минимальных интервалов, содержащих факторы для объектов, в которых основной признак принимает критическое значение. Идея построения таких интервалов взята из технической теории надежности, в которой подобные интервалы называются допусками на параметры работающих изделий. Однако в отличие от задач надежности в данном случае допуски строятся для объектов, у которых основной признак принимает критические значения. Количество арифметических операций, необходимых для определения этих интервалов, пропорционально произведению числа факторов на число объектов обучающей выборки. В результате возникает компактный, быстродействующий и удобный в приложениях алгоритм распознавания образов. Если вектор факторов объекта выборки принадлежит построенному прямоугольнику, то объект относится к критическим, в противном случае – к некритическим. Качество работы прямоугольного решающего правила определяется отношением правильно распознаваемых объектов к общему числу объектов исходной выборки. По построению все критические объекты распознаются правильно, однако некоторые некритические объекты могут ошибочно относиться к критическим. Нетрудно строго доказать, что при расширении набора факторов качество распознавания на основе прямоугольного решающего правила возрастает. Заметим, что качество распознавания в этом случае аналогично коэффициенту множественной корреляции основного признака с набором факторов. Однако процедура вычисления этого показателя значительно проще и удобнее, чем метод главных компонент, используемый в математической статистике для определения коэффициента множественной корреляции.

Удобство использования данного алгоритма позволило провести многочисленные вычислительные эксперименты с накопленными статистическими данными по медицинской экологии, климатологии и морской биологии. Следует подчеркнуть важное обстоятельство при работе с фактическим материалом, в которой взаимодействуют специалист из предметной области (например, медицинской экологии или эпидемиологии) и математик, занимающийся обработкой данных. Во многих известных примерах математики сталкивались с так называемыми трудно формализуемыми системами и привлекались к построению некоторой микротeorии из предметной области. Практика показала, что такой способ взаимодействия представителя предметной области и математика является весьма затратным по времени, усилиям и необходимым ресурсам. Математику зачастую приходится выполнять не свойственные ему функции – сбор дополнительной информации, формулировки содержательных гипотез и преобразование этих гипотез в некоторые уравнения. Причем число неизвестных коэффициентов в этих уравнениях может быть велико, а объем наблюдений для оценки коэффициентов явно недостаточным. Предлагаемая постановка задачи принципиальна, важна в чисто содержательном плане, поскольку позволяет снять проблему «нелинейности». В то же время данный вариант распознавания образов открывает возможность выбирать экспертным способом величину критического уровня для основного признака и тем самым проверять различные содержательные гипотезы о связи основного признака с факторами. Теория катастроф показывает: в окрестности

катастрофы (критического уровня основного признака) его зависимость от факторов упрощается, что свидетельствует в пользу предлагаемого алгоритма распознавания. С другой стороны, многочисленные вычислительные эксперименты, которые удалось провести с реальными фактическими данными благодаря компактности и удобству вычислений по предлагаемому алгоритму показали, что повышение критического уровня, как правило (хотя и не всегда), улучшает качество распознавания.

Особого внимания заслуживает крайне актуальная задача, связанная с возможностью адаптации прямоугольного решающего правила к задачам прогнозирования превышения основным признаком критического уровня. Вычислительные эксперименты показали, что использование прямоугольного решающего правила здесь не всегда дает удовлетворительные результаты. Однако если на конечном отрезке временного ряда выделить точки, где основным признаком превышает критический уровень, и по значениям факторов в предшествующих им точках построить соответствующие интервалы, то для прогнозирования превышения критического уровня основным признаком можно использовать метод голосования факторов. Речь идет о подсчете числа факторов, которые в канун прогнозируемого года попадают в свои интервалы. Если число этих факторов превышает половину, то в прогнозируемом году считается, что основным признаком превысит критический уровень. Для такой модификации прямоугольного решающего правила, во-первых, вычислительная сложность остается прежней, а во-вторых, качество прогнозирования в проведенных объемных численных экспериментах достаточно удовлетворительно. Причем сама процедура модификации решающего правила во многом основывалась на удобствах в его вычислении.

E-mail: cfpd@amur.ru.

УДК 616.72-002.77-085:616.351-091

А.М. Борбат

(Ярославская государственная медицинская академия)

КЛИНИКО-МОРФОЛОГИЧЕСКИЙ АНАЛИЗ ДЛЯ ЦЕЛЕЙ ПОСТРОЕНИЯ ПРОГНОСТИЧЕСКОЙ МОДЕЛИ ЭФФЕКТИВНОСТИ ТЕРАПИИ РЕВМАТОИДНОГО АРТРИТА

Исследована возможность комплексной клинической и морфологической, по результатам исследования биоптатов прямой кишки, оценки для определения прогноза эффективности терапии ревматоидного артрита базисными препаратами до начала лечения.

Ключевые слова: ревматоидный артрит, прогноз, биоптат прямой кишки, базисная терапия.

Ревматоидный артрит – заболевание, характеризующееся значительной длительностью, высокой частотой инвалидизации и сокращением продолжитель-