

А. А. Ченин, И. М. Шабалина, В. А. Кузнецов (Петрозаводск, ПетрГУ). **Модели и методы выявления склонности подростков к употреблению психоактивных средств.**

Основная проблема классификации объектов на основании обучающей выборки, представляющей собой анкетные данные, состоит в разработке алгоритмов, основанных на статистических методах построения диагностических правил. Особенностью анкетных данных является то, что признаки (вопросы) имеют номинальные значения, причем количество градаций разных признаков (ответов на каждый вопрос) может быть разным.

Для решения проблемы несовпадения количества ответов на вопросы анкеты выполняется дихотомизация. Каждую отметку номинальной шкалы считают отдельным самостоятельным признаком, значением которого может быть 0 или 1.

Для сокращения количества признаков необходимо выявить наиболее информативные. Решающим критерием информативности признаков в задаче классификации является величина потерь от ошибок распознавания. Для этого используется метод, основанный на применении логарифма отношения правдоподобия — меры информативности, а так же выявлению зависимых признаков с помощью коэффициентов хи-квадрат и А. А. Чупрова.

Для классификации объектов и отнесения их к одному из заданных классов, требуется построить функцию принадлежности, позволяющую по значениям признаков объекта определить, к какому подмножеству он принадлежит.

Реализованные авторами методы для построения диагностических правил:

Байесовский подход, основанный на вычислении частот возникновения значений признаков и применении формулы Байеса.

Построение линейных разделяющих функций, заключенное в следующем: необходимо построить функцию принадлежности следующего вида:

$$f(x_{i1}, x_{i2}, \dots, x_{im}) = \sum_{k=1}^m w_k x_{ik},$$

где x_{ik} — значение k -го признака у i -го объекта, w_k — весовой коэффициент, $k = 1, 2, \dots, m$, $i = 1, 2, \dots, N$, m — количество обследуемых объектов. На основании функции принадлежности строится диагностическое правило. В качестве весовых коэффициентов используются:

- собственный вектор таблицы сопряженности,
- для признаков используется их прогностические коэффициенты (логарифм отношения правдоподобия),
- энтропия каждого признака.

Полипроцедурный подход на основании информации об уровнях значимости и надежности ранее примененных методов и построенной ими классификации относит объект к одной из групп.

Указанные методы прошли апробацию на анкетах, предоставленных кафедрой психиатрии ПетрГУ и использовались при разработке системы поддержки принятия решений.