

*Н. С. Толстошеин, Д. А. Михин**

**ПРИМЕНЕНИЕ МЕТОДОВ КЛАССИФИКАЦИИ
ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ
В СИСТЕМАХ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ**

В условиях активного роста объема цифровых данных [1], доступной организациям, роль систем поддержки принятия решений (СППР) становится критически важной. СППР призваны трансформировать входные данные в полезные знания, облегчая процесс принятия управленческих решений. Ключевым инструментом в этой трансформации является интеллектуальный анализ данных (в современной научной литературе встречается Data Mining). Интеллектуальный анализ данных применяет современные алгоритмы и методы для обнаружения скрытых закономерностей и тенденций в больших наборах данных, которые зачастую неочевидны для человека.

Интеллектуальный анализ данных – это междисциплинарная область, объединяющая методы из статистики, машинного обучения,

* Работа выполнена под руководством доктора технических наук, профессора кафедры «Информационные системы и защита информации» ФГБОУ ВО «ТГТУ» В. В. Алексева.

теории баз данных и визуализации. Основная цель – извлечение ранее неизвестных, потенциально полезных знаний из данных. Анализ источников [2, 3] позволил сделать вывод, что в контексте СППР, Data Mining выполняет функции, представленные на рис. 1.



Рис. 1. Функции Data Mining в СППР

В процессе анализа данного рисунка был сделан вывод, что центральную роль в задачах интеллектуального анализа данных играет классификация, позволяя автоматически определять принадлежность объектов к заранее заданным классам на основе имеющихся данных.

Анализ источников [4] позволил сделать вывод, что основными методами классификации, применяемыми в Data Mining, являются методы, представленные на рис. 2.

Краткая характеристика методов, представленных на рис. 2:

- методы, основанные на деревьях решений – структура дерева, где каждый внутренний узел представляет проверку некоторого атрибута, каждая ветвь – результат этой проверки, а каждый лиственный узел – присвоение классу;
- байесовские методы основаны на теореме Байеса с сильным предположением о независимости признаков друг от друга, условно говоря, при заданном классе;
- методы, основанные на правилах, строят набор правил вида «ЕСЛИ (условие на атрибуты) ТО (класс)»;

Методы классификации, применяемые в Data Mining



Рис. 2. Методы классификации, применяемые в Data Mining

– методы, основанные на соседях, – класс нового объекта определяется большинством классов его ближайших соседей в обучающей выборке, где «близость» обычно измеряется евклидовым расстоянием;

– методы, основанные на ансамблях, – эти методы объединяют предсказания нескольких базовых моделей для получения более точного и устойчивого результата;

– методы, основанные на разделяющей гиперплоскости, занимают поиском гиперплоскости, которая наилучшим образом разделяет классы в пространстве признаков, максимизируя зазор между классами.

В ходе исследования был поведен анализ вышеперечисленных методов классификации, выявлены их преимущества и недостатки, представленные в табл. 1.

1. Преимущества и недостатки методов классификации Data Mining

Методы	Преимущества	Недостатки
Методы, основанные на деревьях решений	Легкость интерпретации, наглядность, не требуют нормализации данных	Могут быть склонны к переобучению, чувствительны к небольшим изменениям в данных
Байесовские методы	Простота, высокая скорость обучения и предсказания, хорошо работает даже при небольшом объеме обучающей выборки	Предположение о независимости признаков редко выполняется в реальных данных, что может снижать точность
Методы, основанные на правилах	Правила понятны и легко интерпретируются, могут быть извлечены и использованы отдельно	Формирование оптимального набора правил может быть вычислительно сложной задачей
Методы, основанные на соседях	Простота, отсутствие явного этапа обучения, способность моделировать сложные нелинейные границы классов	Чувствительность к шкале признаков, высокая вычислительная стоимость на этапе предсказания для больших наборов данных
Методы, основанные на ансамблях	Очень мощный и гибкий метод	Более сложен в настройке, чувствителен к параметрам, может переобучаться при неправильной настройке
Методы, основанные на разделяющей гиперплоскости	Хорошо работает в пространствах высокой размерности, эффективен, когда классы четко разделимы	Чувствителен к выбору ядра и его параметров, сложен в интерпретации, медленный на больших наборах данных, плохо работает с зашумленными данными

В результате анализа таблицы был сделан вывод, что выбор метода классификации зависит от количества объектов и признаков, наличия пропущенных значений, типов признаков, сложности задачи, необходимой прозрачности модели и располагаемых вычислительных ресурсов.

Часто для решения реальных задач интеллектуального анализа данных пробуют несколько разных методов и выбирают тот, который дает наилучшие результаты на задаче, что обычно определяется с помощью кросс-валидации.

Несмотря на огромный потенциал, применение классификации в СППР обладает рядом ограничений:

- низкое качество данных приведет к неточным моделям;
- сложные модели могут быть «черными ящиками», что затрудняет понимание логики принятия решений;
- статистические свойства данных могут меняться со временем, требуя регулярного переобучения моделей.

Таким образом интеграция методов классификации из области Data Mining в системы поддержки принятия решений открывает новые возможности для современных организаций.

Разнообразие алгоритмов классификации позволяет подобрать наиболее подходящее решение для конкретной задачи. Каждый метод обладает своими преимуществами и недостатками, и выбор оптимального подхода часто требует экспериментов и учета специфики предметной области.

Несмотря на существующие ограничения, дальнейшее развитие технологий машинного обучения, повышение доступности вычислительных ресурсов и растущее понимание важности данных приведут к тому, что методы классификации будут играть все более значимую роль в эволюции СППР.

Список литературы

1. Дмитриев, А. П. Стремительный рост цифровых данных: анализ мировых трендов и прогноз развития в России / А. П. Дмитриев, С. Ш. Лейба // Региональная и отраслевая экономика. – 2024. – № 1. – С. 141 – 152.
2. Арсентьева, Н. В. Data mining: методы, этапы, применение и значение в современном мире / Н. В. Арсентьева, А. А. Скрипин, И. И. Скрипина // Форум молодых ученых. – 2024. – № 6(94). – С. 97 – 101.

3. Алексеев, В. В. Системный подход к организации структуры управления мобильной сетью в интересах обеспечения устойчивости связи / В. В. Алексеев, Д. А. Иванов, И. Г. Рыжов // Вестник Тамбовского государственного технического университета. – 2025. – Т. 31, № 2. – С. 259 – 268.

4. Шабанов К. Б. Применение методов интеллектуального анализа данных для повышения качества принятия решений при управлении ресурсами информационной медиасистемы / К. Б. Шабанов, В. В. Алексеев // Вестник Тамбовского государственного технического университета. – 2021. – Т. 27, № 1. – С. 14 – 19.

*Кафедра «Информационные системы и защита информации»
ФГБОУ ВО «ТГТУ»*