

ОСОБЕННОСТИ ПРИМЕНЕНИЯ ТЕХНОЛОГИЙ DATAMINING

© 2021 Ю. П. Преображенский, О. Н. Чопоров, Е. Ружицкий

*Воронежский институт высоких технологий (Воронеж, Россия)
Воронежский государственный технический университет (Воронеж, Россия)
Панъевропейский университет (Братислава, Словакия)*

В данной работе рассматриваются особенности обработки данных в информационных системах.

Ключевые слова: обработка данных, информационная система, модель.

Экспериментальные данные во многих случаях могут быть представлены на основе компьютерных форматов. Необходимо обеспечивать процедуры, связанные с интерпретацией, анализом и обработкой. Что можно достичь с их помощью?

Есть возможности для того, чтобы продемонстрировать те закономерности, которые будут потенциально возможными. Также будет прогнозирование новых фактов.

В результате мы можем оценивать значение в целевом признаке Y при рассмотрении разных объектов a , когда есть описание x , которое анализируется в виде множества независимых переменных. Существуют разные сопутствующие проблемы в ходе многомерного [1, 2] анализа систем.

Например, требуется осуществлять сравнение анализируемых объектов с точки зрения их идентичности. Отобранные объекты должны быть правильным образом отклассифицированы. Важно увидеть причинно-следственные связи среди объектов.

В течение достаточно долгого времени оценка по Р. Фишеру рассматривалась в виде характеристики, демонстрирующей особенности взаимодействий в экспериментах [3, 4].

Современные практические исследования требуют внедрения комплексных подходов. В них таблицы размещаются концентрированным образом внутри хранилищ данных.

Мы предоставляем разрозненную информацию как многомерный куб. есть возможности для манипулирования им. Чтобы сложные гипотезы были проверены, а также стратегические проблемы были разрешены, применяется аппарат, связанный с тем, как происходит извлечение знаний из различных хранилищ [5, 6].

Процессы, реализуемые в ходе технологии Data Mining, рассматриваются в виде аналога извлечения ценных веществ из добытых материалов. Научные гипотезы извлекаются на этапе Data Mining. На этапе Data Analysis будет осуществляться дальнейший анализ. Они рассматриваются в виде двух непрерывных комплексных процессов.

Модели в окружающем мире являются принципиально множественными. Результаты, связанные с пассивными наблюдениями, активным образом применяются в ходе современных исследований.

Весьма трудно осуществить постановку рандомизированного управляемого эксперимента. То есть, важно при том, что мы, не зная всех влияющих параметров, смогли определить возможные конкурирующие гипотезы. Для произвольного набора данных трудно выбрать идеальный метод.

Существует так называемый принцип множественности моделей. В мультимодальном выводе исходят из того, что параметры будут оцениваться с привлечением не одной модели [7, 8].

Моделей будет много. Качество входных данных будет оказывать влияние на возможности применения статистического подхода.

Исследователям, когда они проводят разработку статистических моделей, необходимо разрабатывать алгоритмы, позволяющие извлекать закономерности. Для их прогнозирования могут быть применены

Преображенский Юрий Петрович – Воронежский институт высоких технологий, канд. техн. наук, профессор, Petrovich@vvt.ru.

Чопоров Олег Николаевич – Воронежский государственный технический университет, профессор, choporov_oleg@mail.ru.

Ружицкий Евгений – Панъевропейский университет, канд. техн. наук, доцент, rush_ev_g_br53@yandex.ru.

разные подходы. Например, используется технология бустинга, искусственных нейронных сетей, подход, в котором есть «случайный лес».

Метод может выбираться исходя из того, какой будет опыт у исследователей. Иногда исследователи могут не иметь информацию относительно существующих моделей. В этой связи важно осуществлять сравнительную диагностику относительно большого числа моделей. Интересно с точки зрения практики то, что в статистическом анализе, машинном обучении, технологии Data Mining применяются одинаковые алгоритмы.

Когда происходит решение разных практических задач, тогда обучение может базироваться на том, что привлекается учитель или процесс может осуществляться без него [9, 10]. Кластеризацию и ординацию привлекают при поиске закономерностей, если нет учителя.

Принцип дискретности применяется в кластеризации. Тогда исследователи стремятся к определению оптимального разбиения исходных совокупностей относительно отдельных групп, сформированных на базе однородных объектов. Должно быть максимальное возможное различие среди групп.

Принцип континуальности лежит в основе ординации.

Во многих моделях, которые привлекают обучение с учителем выделяют функцию, связывающую параметры функции и аргументы.

За счет варьирования функций есть возможности для выбора моделей. Также есть влияние вида обрабатываемых данных на выбор подходящего семейства моделей. Будет влиять и закон распределения данных. То, насколько достигнутые результаты будут адекватными, определяется типом переменных, которые будут составлять таблицу предикторов.

Нет необходимости в том, чтобы проводить анализ структуры внутренних взаимосвязей среди предикторов и в ходе прогнозирования. Основной задачей считается предсказание величины целевого признака, исходя из вариации по значению признаков. Если для моделируемых процессов мы видим достаточно сложные закономерности, тогда подобные модели могут характеризоваться очень большим числом параметров. Чтобы их построить, необходимо опираться на соответствующие принципы.

Первый из них связан с выделением таких моделей, которые могут осуществлять

последовательные переходы от вариантов, являющихся простыми, к таким, которые будут весьма сложными.

Второй исходит из особенностей развития моделей. Модели прогнозов при том, что поддерживается необходимая точность в предсказании, могут заметным образом ее потерять с точки зрения интерпретируемости. Не во всех случаях легко осуществить предметный анализ по нескольким десяткам моделей, по их коэффициентам. То есть, регрессионные подходы в подобных случаях трудно реализовать.

Это связано с природой многомерных откликов. Задача, связанная с формированием статистических моделей, относится к тому, что объясняется изменчивость в многомерных откликах.

Такие отклики представляются как замкнутая система элементов, которые относятся к множеству, содержащем разные типы элементов. Происходит разбиение элементов по классам на базе классификатора [11, 12].

Матрицы случайных наблюдений, связанные с многомерными откликами, формируются при помощи композиций по разным объектам. Важно осуществлять оценку зависимостей таких матриц от совокупности экзогенных переменных, являющихся независимости.

Подобные модели могут быть сформированы при помощи двух способов. В первом способе осуществляется процесс свертки многомерных откликов к одномерным. Для этого применяется соответствующая функция. Затем будут применяться регрессионные модели. Как результат можно прийти к потере части информации.

Корреляционная матрица с точки зрения структуры может меняться, если применяется многомерный анализ.

ЛИТЕРАТУРА

1. Голованчиков А. Б. Сравнение точности аппроксимации экспериментальных данных методом наименьших квадратов с методом наименьших квадратов / А. Б. Голованчиков, М. К. Доан, А. В. Петрухин, Н. А. Меренцов // Моделирование, оптимизация и информационные технологии. – 2020. – Т. 8. – № 1 (28). – С. 38-39.
2. Львович Я. Е. Адаптивное управление марковскими процессами в конфликтной ситуации / Я. Е. Львович, Ю. П. Преображенский, Р. Ю. Паневин // Вестник Воро-

нежского государственного технического университета. – 2008. – Т. 4. – № 11. – С. 170-171.

3. Воронов А. А. Обеспечение системы управления рисками при возникновении угроз информационной безопасности / А. А. Воронов, И. Я. Львович, Ю. П. Преображенский, В. А. Воронов // Информация и безопасность. – 2006. – Т. 9. – № 2. – С. 8-11.

4. Калаев В. Н. Регрессионный анализ в биологических исследованиях / В. Н. Калаев, Е. А. Калаева, А. П. Преображенский, О. В. Хорсева // Системный анализ и управление в биомедицинских системах. – 2007. – Т. 6. – № 3. – С. 755-759.

5. Львович И. Я. Основы информатики / И. Я. Львович, Ю. П. Преображенский, В. В. Ермолова. – Воронеж, 2014. – 339 с.

6. Черников С. Ю. Использование системного анализа при управлении организациями / С. Ю. Черников, Р. В. Корольков // Моделирование, оптимизация и информационные технологии. – 2014. – № 2 (5). – С. 16.

7. Чопоров О. Н. Методика преобразования качественных характеристик в численные оценки при обработке результатов медико-социального исследования / О. Н. Чопоров, А. И. Агарков, Л. А. Куташова, Е. Ю. Коновалова // Вестник Воронежского института высоких технологий. – 2012. – № 9. – С. 96-98.

8. Преображенский А. П. Методы прогнозирования характеристик рассеяния элек-

тромагнитных волн / А. П. Преображенский // Моделирование, оптимизация и информационные технологии. – 2014. – № 1 (4). – С. 3.

9. Львович Я. Е. Оптимизация проектирования многоаспектной цифровой среды системы однородных объектов на основе процедур декомпозиции и агрегации / Я. Е. Львович, А. В. Питолин, С. О. Сорокин // Моделирование, оптимизация и информационные технологии. – 2019. – Т. 7. – № 2 (25). – С. 186-195.

10. Фролов С. В. Двухэтапная процедура синтеза управления нелинейными нестационарными объектами с использованием многослойного персептрона / С. В. Фролов, С. В. Синдеев, А. А. Коробов, К. С. Савинова, А. Ю. Потлов // Моделирование, оптимизация и информационные технологии. – 2020. – Т. 8. – № 3 (30). – С. 12-13.

11. Сай В. К. Классификационный подход на основе комбинации глубоких нейронных сетей для прогнозирования отказов сложных многообъектных систем / В. К. Сай, М. В. Щербаков // Моделирование, оптимизация и информационные технологии. – 2020. – Т. 8. – № 2 (29). – С. 33-34

12. Ломакина Л. С. Классификация потоковых данных на основе байесовского критерия / Л. С. Ломакина, А. Н. Субботин // Моделирование, оптимизация и информационные технологии. – 2020. – Т. 8. – № 1 (28). – С. 18-19.

THE FEATURES OF APPLICATION OF TECHNOLOGIES DATAMINING

© 2021 Yu. P. Preobrazhenskiy, O. N. Choporov, E. Ruzhicky

Voronezh Institute of High Technologies (Voronezh, Russia)

Voronezh State Technical University (Voronezh, Russia)

Pan-European University (Bratislava, Slovakia)

This paper discusses the features of data processing in information systems.

Keywords: data processing, information system, model.