

22 Кластер-анализ

Что такое анализ кластерный – один из математических методов, заключающийся в том, что определенный набор объектов разбивают на группы, которые называются кластерами.

Кластерный анализ предназначен для разбиения совокупности объектов на однородные группы (кластеры или классы). Это задача многомерной классификации данных.

Существует около 100 разных алгоритмов кластеризации, однако, наиболее часто используемые:

иерархический кластерный анализ,
кластеризация методом k-средних.

Где применяется кластерный анализ:

1. В маркетинге это сегментация конкурентов и потребителей.
2. В менеджменте:
 - разбиение персонала на различные по уровню мотивации группы, классификация поставщиков,
 - выявление схожих производственных ситуаций, при которых возникает брак.
3. В медицине — классификация симптомов, пациентов, препаратов.
4. В социологии — разбиение респондентов на однородные группы.

По сути кластерный анализ хорошо зарекомендовал себя во всех сферах жизнедеятельности человека. Прелесть данного метода — он работает даже тогда, когда данных мало и не выполняются требования нормальности распределений случайных величин и другие требования классических методов статистического анализа.

Кластерный анализ исполняет такие главные задачи:

1. Изучение концептуальных полезных схем группирования объектов.
2. Разработка классификации или типологии. Порождение гипотез на основании исследования данных.

3. Проверка исследования или гипотез для определения, действительно ли группы (типы), выделенные каким-либо методом, есть в имеющихся данных.

Вне зависимости от предмета изучения использование кластерного анализа предусматривает следующие стадии:

1. Отбор выборки для кластеризации. Понимается, что есть смысл кластеризовать лишь количественные данные.

2. Определение переменных, по которым будут оценивать объекты в выборке, то есть признакововое пространство.

3. Вычисление значений определенной меры различия или сходства меж объектами.

4. Использование способа кластерного анализа для того, чтобы создать группы сходных объектов.

5. Проверка достоверности итогов кластерного решения.

23 Планирование эксперимента

Планирование эксперимента - это процесс выбора условий, процедуры и методов проведения опытов, их числа и условий, необходимых и достаточных для решения поставленной задачи с требуемой точностью.

Требования к планированию эксперимента:

- 1) число опытов должно быть минимальным, чтобы не усложнять процедуру эксперимента и не увеличивать его стоимость, но не в ущерб точности результата;
- 2) необходимо определить совокупность факторов, влияющих на результаты эксперимента, ранжировать их, выявить главные, а несущественные переменные можно исключить;
- 3) условием корректности эксперимента следует считать одновременное варьирование всеми переменными (факторами), оказывающими взаимное влияние на исследуемый процесс;
- 4) ряд действий в эксперименте может быть заменен их моделями (прежде всего математическими), при этом адекватность моделей должна быть проверена и оценена;
- 5) необходимо разработать стратегию эксперимента и алгоритм ее реализации: серии эксперимента должны анализироваться после завершения каждой из них перед переходом к последующей серии.

План проведения эксперимента должен включать следующие разделы:

- 1. Наименование темы исследования.
- 2. Цель и задачи эксперимента.
- 3. Условия проведения эксперимента: параметр оптимизации и варьируемые факторы.
- 4. Методика проведения исследования.
- 5. Обоснование количества опытов (объема эксперимента).
- 6. Средства и методика проведения измерений.
- 7. Материальное обеспечение эксперимента (перечень оборудования).
- 8. Методика обработки и анализа экспериментальных данных.
- 9. Календарный план проведения испытаний, в котором указываются сроки их выполнения, исполнители, представляемые данные эксперимента.
- 10. Смета расходов.

Цель и задачи эксперимента - исходный пункт плана. Они формулируются на основе анализа научной гипотезы, теоретических результатов собственного исследования либо исследований других авторов.

Цель определяет конечный результат эксперимента, т. е. то, что исследователь должен получить в итоге. Например, подтвердить правильные научные гипотезы; проверить на практике адекватность, работоспособность и практическую пригодность моделей, методик; определить оптимальные условия технологического процесса и т. п.

Частными задачами эксперимента при его планировании могут быть:

- - проверка теоретических положений с целью подтверждения их истинности;
- - проверка (уточнение) констант математических либо иных моделей;
- - поиск оптимальных (допустимых) условий какого-либо процесса;
- - построение интерполяционных аналитических зависимостей.

24 Область применения бизнес-анализа

Бизнес-анализ — это инструмент, позволяющий выявлять потребности бизнеса и находить пути преодоления трудностей. Он охватывает широчайший спектр вопросов. Анализ бизнес-процессов следует выполнять не только для начинающих проектов, но и действующих, в которых он существенно повысит вероятность успеха.

Основные задачи бизнес-анализа

1. предоставлять достоверные сведения в нужном разрезе для вынесения решений по управлению;
2. определять, насколько в данный момент эффективны бизнес-процессы;
3. оценивать потребности бизнеса, текущие и стратегические требования компании к эффективной организации бизнес-анализа;
4. формировать стратегию развития предприятия в целом и его самостоятельных направлений;
5. намечать пути реализации целей;
6. оценивать, нужно ли корректировать внутренние и внешние процессы;
7. оценивать и контролировать риски развития; обеспечивать координированную работу всех отделов.

Сфера применения бизнес анализа весьма обширна – это и разработка концепции IT решения, и описание и реинжиниринг бизнес процессов автоматизируемых предприятий, и проектирование информационных систем, а также анализ, детализация и управление требованиями на всех этапах разработки.

Основная роль аналитика на проекте – это проанализировать и систематизировать информацию и обеспечить единое понимание содержания проекта у всех задействованных в проекте лиц.

25 Стохастическая связь данных

Стохастическая связь – это связь между величинами, при которой одна из них, случайная величина y , реагирует на изменение другой величины x или других величин x_1, x_2, \dots, x_n (случайных или неслучайных) изменением закона распределения. Это обуславливается тем, что зависимая переменная (результативный признак), кроме рассматриваемых

независимых, подвержена влиянию ряда неучтенных или неконтролируемых (случайных) факторов, а также некоторых неизбежных ошибок измерения переменных. Поскольку значения зависимой переменной подвержены случайному разбросу, они не могут быть предсказаны с достаточной точностью, а только указаны с определенной вероятностью.

Характерной особенностью стохастических связей является то, что они проявляются во всей совокупности, а не в каждой ее единице (причем неизвестен ни полный перечень факторов, определяющих значение результативного признака, ни точный механизм их функционирования и взаимодействия с результативным признаком). Всегда имеет место влияние случайного. Появляющиеся различные значения зависимой переменной – реализации случайной величины.

Для определения наличия и направления стохастической зависимости между показателями используются такие способы анализа, как аналитические группировки, сравнение параллельных временных рядов данных, аналитические группировки, построение корреляционных полей или диаграмм разброса. Все перечисленные методы позволяют визуально определить наличие связи.

26 Методы аналитической обработки данных в режиме реального времени (OLAP)

OLAP (On-Line Analytical Processing) – это класс приложений и технологий, предназначенных для оперативной аналитической обработки многомерных данных (сбор, хранение, анализ) для анализа деятельности корпорации и прогнозирования будущего состояния с целью поддержки принятия управленческих решений. Технология OLAP применяется, чтобы упростить работу с многоцелевыми накопленными данными о деятельности корпорации в прошлом и не погрязнуть в их большом объеме, а также превратить набор количественных показателей в качественные, позволяет аналитикам, менеджерам и управляющим сформировать свое собственное видение данных, используя быстрый, единообразный, оперативный доступ к разнообразным формам представления информации. Такие формы,

полученные на основании первичных данных, позволяют пользователю сформировать полноценное представление о деятельности предприятия

Функциональность OLAP, как систем реализующих интеллектуальный анализ данных, заключается в динамическом многомерном анализе консолидированных данных предприятия, направленном на поддержание следующих аналитических и навигационных видов деятельности пользователя:

- вычисления и моделирование, примененные к измерениям и/или их конкретным элементам, использующие информацию об иерархиях;
- анализ временных тенденций и взаимозависимостей показателей (анализ трендов), учет которых помогает повысить качество принимаемых оперативных и стратегических решений;
- формирование срезов многомерного представления для просмотра на экране;
- переход к более глубоким уровням детализации;
- доступ к исходным данным;
- «вращение» многомерных представлений: перемещение измерений с целью формирования различных форм представления данных на экране компьютера

OLAP (On-Line Analytical Processing) – это не отдельно взятый программный продукт, не язык программирования и даже не конкретная технология, это совокупность концепций, принципов и требований, лежащих в основе программных продуктов, облегчающих аналитикам доступ к данным. Термин OLAP очень популярен в настоящее время и OLAP-системой зачастую, но не совсем верно, называют любую DSS-систему, основанную на концепции ХД и обеспечивающих малое время выполнения (On-Line) аналитических запросов, не зависимо от того, используется ли многомерный анализ данных.

Применение OLAP – систем. На данное время разработан довольно много аналитических систем, сконструированных с использованием OLAP-технологии (Hyperion OLAP, Elite OLAP, Oracle Express и много других). Рынок программных OLAP-продуктов постоянно расширяется. Современные системы оперативной аналитической обработки дают пользователям возможность решать ключевые задачи управления бизнесом-процессом, в частности прикладные программы Hyperion OLAP разрешают выполнять анализ прибыльности; анализ направлений развития продукции; анализ

продажи; анализ положения на рынке; анализ ассортимента продуктов; анализ риска; анализ конкурентоспособности; складывания отчетов из производительности; моделирования сценария; анализ бюджета и прогнозов и т.п.

27 Многомерные кубы OLAP

Кубы данных OLAP (Online Analytical Processing — оперативный анализ данных) позволяют эффективно извлекать и анализировать многомерные данные. В отличие от других типов баз данных, базы данных OLAP разработаны специально для аналитической обработки и быстрого извлечения из них всевозможных наборов данных. На самом деле существует несколько ключевых различий между стандартными реляционными базами данных, такими как Access или SQL Server, и базами данных OLAP.

OLAP предоставляет удобные быстродействующие средства доступа, просмотра и анализа деловой информации. Пользователь получает естественную, интуитивно понятную **модель данных, организуя их в виде многомерных кубов (Cubes)**. Осями многомерной системы координат служат основные *атрибуты* анализируемого бизнес-процесса. Например, для продаж это могут быть товар, регион, тип покупателя. В качестве одного из *измерений* используется время. На пересечениях осей *измерений* (Dimensions) находятся данные, количественно характеризующие процесс — *меры* (Measures). Это могут быть объемы продаж в штуках или в денежном выражении, остатки на складе, издержки и т. п. Пользователь, анализирующий информацию, может "разрезать" куб по разным направлениям, получать сводные (например, по годам) или, наоборот, детальные (по неделям) сведения и осуществлять прочие манипуляции, которые ему придут в голову в процессе анализа.

28 Виды аналитической обработки данных

аналитическая обработка — технология обработки данных, заключающаяся в подготовке суммарной (агрегированной) информации на основе больших массивов данных, структурированных по многомерному принципу.

Виды аналитической обработки данных

- Хранилища данных
- Реляционные хранилища данных
- Таблица фактов
- Таблицы измерений
- OLAP и многомерные хранилища данных

- Data Mining
- Регрессионный, дисперсионный и корреляционный анализ
- Методы анализа в конкретной предметной области
- Метод "ближайшего соседа"
- Кластерные модели
- Алгоритмы ограниченного перебора
- Эволюционное программирование

3. АНАЛИТИЧЕСКАЯ ОБРАБОТКА И ИЗУЧЕНИЕ ИНФОРМАЦИИ ДЛЯ АНАЛИЗА

Цель аналитической обработки данных

выявление связи между показателями
и установление количественной
характеристики влияния факторов на
тот или иной показатель

Этапы аналитической обработки информации

- выявление связи между показателями
- установление количественной характеристики влияния факторов на тот или иной показатель
- анализ причин изменений и отклонений от плана