ТиИИт лекции

**Модель**

Моделирование можно рассматривать, как замещение исследуемого объекта (оригинала) его условным образом, описанием или другим объектом – моделью.

**Основная задача моделирования**: познание свойств оригинала, через исследование модели.

Чем сложнее объект, тем важнее роль моделирования в его изучении в его изучении и создании.

Реальная польза от моделей может быть получена при выполнении главных условий:

Модель должна быть **адекватной**: точно отображать исследуемые характеристики оригинала.

Модель должна устранять проблемы с физическим измерением характеристик оригинала (т. е. достаточной).

**Виды моделей:**

1. Физические модели.

Реальное воплощение физических свойств оригинала. Часто называют макетами (меньшие масштабы реального объекта).

1. Математические модели

Представляют собой формализованное описание объекта или системы с помощью некоторого языка, например совокупности математических соотношений и формул (или схемы алгоритма). Стохастическая (вероятностная) модель

1. Имитационная модель

Сложная система представляется в виде алгоритма или логических действий (блоков), в которых также используется аппарат математического моделирования, статистики и теории вероятности.

1. Модель ИИ. Это “архитектура” (НС, Решающее дерево, Регрессии, Кластеры, Композиции, Ансамбли) реализующая интересующие характеристики объекта (процесса), параметры которой настроены методами машинного обучения.

По своей структуре они могут напоминать математические модели (например, линейная регрессия)

**Свойства моделей**:

1. Адекватность – это степень соответствия модели исследуемому объекту.

На практике модель считают адекватной, если она с удовлетворительной точностью позволяет достичь целей исследования.

Для настройки модели применяют верификацию. //Для ИИ-обучения.

2)Простота. Чем большее количество свойств объекта описывает модель, тем выше ее адекватность (или полезность). Надо стремиться найти более простую модель, позволяющую достичь требуемых результатов.

**Цели моделирования:**

1. Изучение объекта – познавательная цель. Механизмов функционирования, свойств и характеристик (и как следствие – прогнозирование состояний).
2. Управление объектами и системами с целью выработки оптимальных управляющих воздействий и характеристик системы.

(Адаптивное управление/настройка)

**Комплексное моделирование**

1. Определение цели.
2. Разработка концептуальной модели.
3. Формализация модели.
4. Программная реализация.
5. Планирование модельных экспериментов.
6. Реализация плана экспериментов.
7. Анализ и интерпретация результатов.

**Погрешности моделирования**

Точное аналитическое решение удается получить редко, поэтому каждый результат необходимо учитывать с погрешностью.

Ошибка: error = x -

Абсолютная погрешность: e = |x - |

Относительная ошибка:

Источники погрешности:

Применяемые для решения задач методы в большинстве случаев являются приближенными (погрешность метода)

При вводе данных в ЭВМ и выполнении операций производятся округления (вычислительная погрешность)

1) Модель является приближенным описанием реального процесса (погрешность модели)

2) Исходные данные содержат погрешность, т. к. являются результатом измерений (погрешность)

3) Применяемые для решения задач методы в большинстве случаев являются приближенными (погрешность метода)

4) При вводе данных в ЭВМ и выполнении операций производятся округления (вычислительная погрешность)

**Вычислительные методы в моделировании:**

1. Метод эквивалентных преобразований. Исходную задачу заменяют другой, имеющей то же решение.

Пример: найти корень уравнения f(x) = 0 Ф(x) =

Можно заменить задачей поиска глобального минимума

1. Метод аппроксимации. Замена исходной задачи другой, решение которой близко к решению исходной задачи.
2. Конечно-разностные методы. Замена производных конечными разностями.

Например:

1. Прямые методы – решение получается за конечное число элементарных операций (последовательные вычисления).
2. Итерационные методы – методы последовательных приближений к решению задачи.

Пример: Вычисление определенного интеграла, поиск корней уравнения (методы Ньютона, Хорд и т. д.)

1. Методы статистических испытаний (Монте-Карло)-основаны на моделировании случайных величин и построении статистических оценок решения задач.

Для реализации решений используют генераторы случайных чисел.

RNG – random numbers generator.

**Имитационное моделирование**

**Задачи моделирования**

Для каждой моделируемой ситуации известна цель (или несколько целей), достижение которой (которых) считается желательным.

В прикладных исследованиях приходится иметь дело со сложными системами, в которых имеется множество целевых функций, и количественное выражение этих функций задача затруднительная.

Основные сферы применения компьютерного моделирования:

-решение оптимизационных задач -исследование сложных систем -прогнозирование их будущих состояний в зависимости от избираемых стратегий управления.

**Общие свойства имитационного моделирования**

Процесс функционирования сложной системы представляется в виде определенного алгоритма, то есть логических действий, которые и реализуется на компьютере. По результатам реализации могут быть сделаны те или иные выводы относительно исходного процесса.

Имитационное моделирование предоставляет возможность преодолеть аналитические трудности и найти ответ на поставленные вопросы о поведении системы.

**Этапы создания имитационной модели:**

1. Формулирует вопросы о поведении сложной системы.
2. Декомпозиция системы.
3. Формулируют Гипотезы (Н).
4. Задание системного времени.
5. Задание свойств системы.
6. Случайные параметры заменяются машинной реализацией.

**Характеристики имитационного моделирования:**

Имитационная модель – алгоритм, имитирующий реальный объект с учетом случайных входных воздействий и внешней среды.

Имитационная модель является объектом эксперимента.

Имитационная модель – стохастический процесс смены дискретных состояний системы. При реализации модели производится накопление статистических данных по показателям модели, которые являются предметом исследований. По окончании моделирования накопленная статистика обрабатывается, и результаты моделирования получаются в виде выборочных распределений исследуемых величин. Таким образом, математическая статистика и теория вероятностей являются математическими основами имитационного моделирования.

**Корреляция**

Коэфф. Корреляции двух величин X и Y

Корреляция Пирсона:

С помощью коэффициента корреляции Пирсона можно определить силу линейной зависимости между величинами. Для независимых r = 0 (но не наоборот).

Сам по себе факт корреляционной зависимости не дает основания утверждать, что одна из переменных предшествует или является причиной изменений, или то, что переменные вообще причинно связаны между собой.

Статистика: <частное> - <общее>

**Генерирование случайных чисел**

Для метода статического моделирования на ЭВМ характерно то, что большое число операций, соответственно и большая доля машинного времени расходуется на действия со случайными числами.

Результаты статического моделирования существенно зависят от качества исходных последовательностей случайных чисел. Поэтому наличие простых и экономических способов формирование последовательностей случайных чисел требуемого качества во многом определяет возможности практического использования машинного моделирования систем.

При моделировании систем на ЭВМ программная имитация случайных воздействий любой сложности сводится к генерированию некоторых стандартных (базовых) случайных процессов и к их последующему функциональному преобразованию.

Как показывает практика, оптимальным базовым процессом является представляющих собой реализацию равномерно распределенной на интервале (0, 1)

Методы серединных квадратов.

1. Выбирается число разрядов n – целое значение.
2. Задается 2n-разрядное число x0 = a1, a2,…a2n, меньше 1.
3. Текущее число возводится в квадрат. (x0)^2 = b1, b2,…..b4n
4. Определяется новое значение искомого числа x1 = bn+1, bn+2…..b3n, путем выделения 2n средних разрядов из квадрата исходного числа.
5. x0 = x1, переход к пункту 3.

Рекурсивный метод.

Zi = a0 \* Zi-2 + ai\*Zi-1

Основан на формуле:

Xi = Zi mod M

Где а0, а1 – целые положительные коэффициенты;

М – целое положительное число, модуль;

Zi – рассчитываемые целые положительные числа;

xi – рассчитываемые квазиравномерные числа.

Нормальное распределение.

**Механизмы принятия решений**

Процессы принятия решений в различных сферах деятельности во многом аналогичны. Поэтому необходим универсальный метод поддержки принятия решений.

ИС: DSS, OLAP

Технологии:DM, KDD

- Зачастую перед ЛПР (лицо принимающее решение) стоит задача выбора между различными альтернативами.

- ЛПР руководствуется своим опытом и интуицией (неопределенный характер решений, субъективное решение).

- С целью придания ясности процесс подготовки принятия решения сопровождается количественным выражением категорий («важнее»)

Имеются:

1. несколько альтернатив (объектов, действий и т. п.),
2. главный критерий (главная цель) сравнения альтернатив (функционал для оптимизации).
3. Несколько групп факторов (частных критериев, объектов, действий и т. п.), влияющих на отбор альтернатив.

f(w, x) = y` ; f – альтернатива

Q (y`) -> opt; Q – оценка альтернативы

Требуется каждой альтернативе поставить в соответствие приоритет – получить рейтинг альтернатив.

Причем чем более предпочтительна альтернатива по избранному критерию, тем больше ее приоритет.

Типовая задача принятия решения:

1. Рассматриваются несколько вариантов решения,
2. Задан критерий, по которому определяется в какой мере то или иное решение является подходящим,

Виды задач принятия решений

1. Задача выбора.

Выбрать или отвергнуть несколько вариантов из группы возможных

1. Задача распределения ресурсов.

- Множество вариантов

**Типовые задачи:**

1. Работа с клиентами.

Какой из клиентов чаще покупает мои товары? Кто из потенциальных клиентов является наиболее перспективным?

Планирование рекламных компаний?

Рекомендательные системы

CRM-системы для сбора данных

1. Анализ рисков.

Анализ рисков инвестирования, составление портфеля активов.

Анализ в Страховании.

Выдача кредитов.

1. Распределение ресурсов

Пример. Руководство предприятия рассматривает перспективные проекты развития. Для них создается рейтинговая модель. В итоге каждому проекту приписывается доля от единицы. Эти доли показывают, какой процент от имеющихся ресурсов надо вложить в каждый проект.

1. Планирование от достигнутого.

Исходя из имеющихся: основных фондов кадров, сырья, инфраструктуры, партнеров, конкурентов и пр. составляется рейтинг возможных положений предприятия через год (месяц, квартал).

Моделирование в стрессовых ситуациях.

1. Планирование

- Анализ перспективных стратегий развития.

6. Анализ эффективность-стоимость. (цена-качество)

Выбор направления деятельности (затрат) с максимальной отдачей при минимальных потерях.

7.Поиск существенных факторов.

Задача определения существенных факторов особенно актуальна при решении масштабных проблем и проблем и проблем стратегического планирования.

8.Диагностика.

Анализ текущего состояния объекта.

9.Построение зависимостей.

Комплексная задача, основная на поиске существенных факторов.

Попытка расширить область задачи и данных.

(Рекомендательные системы)

**Сопутствующие задачи**

1. Выявить наиболее неясные и противоречивые этапы создания модели для поддержки принятия решения

- является ли рассматриваемый набор решений полным

-учтены ли все группы факторов, влияющих на выбор наиболее приоритетного решения.

-известны ли сравнительные оценки того, как сильно влияют главный критерий, факторы и альтернативы друг на друга, имеются ли противоречия в оценках.

-имеются ли альтернативные мнения по рассматриваемой проблеме.

2) Разбить большую задачу о принятии решения на ряд малых задач.

Позволяет распределить работу по подготовке принятия решения. Представить в понятной форме схему взаимодействия факторов, влияющих на формирование приоритетов решений, и самих решений.

3)Оценить и минимизировать противоречивость данных, использующихся для определения приоритетов рассматриваемых решений (очистка данных)

4)Установить предварительно условия, при которых по найденному рейтингу приоритетов возможных решений можно сделать выбор лучшего решения.

5)Проверить факторы и варианты решений на избыточность (отбор параметров – feature selection, отбор стратегий).

6)Оценить устойчивость результатов, полученных в результате применения метода.

Невозможно гарантировать что данные и модели будут абсолютно точными.

Как изменения в параметрах повлияет на изменение приоритета выбора решения (устойчивость).

**Сопутствующие проблемы**

1. Полный набор решений неизвестен (и факторов).

Так же факторы могут иметь свою динамику: С течением времени малозначительные на данный момент факторы и альтернативы могут стать важнейшими, и наоборот, то что важно сейчас может оказаться незначительным впоследствии.

1. Большой набор альтернатив и факторов, влияющих на их отбор.

Необходимы способы разбиения совокупности рассматриваемых решений и факторов, которые потенциально определяют приоритеты решений, на достаточно малые группы – кластеры.

1. В реальных задачах влияния различных факторов на выбор оптимального решения сложны и запутаны.

- нелинейные зависимости

- рекурсивные зависимости (факторы, влияющие на решение, сами зависят от решения)

4) Нет точной количественной информации, необходимой для решения задачи.

- проблема добычи данных

5) Имеющиеся данные противоречивы.

- поиск противоречий

- устранение противоречий

6) Нет четкой и универсальной методики составления рейтинга рассматриваемых решений.

- эксперты предлагают разные системы оценок, оценки разнятся.

- как сформировать целевую функцию (Q -> min)?

Для поддержки принятия решений с помощью IT, включая анализ и выработку альтернатив, в СППР используются следующие технологии: информационный поиск, интеллектуальный анализ данных, извлечение знаний в базах данных, рассуждение на основе прецедентов, имитационное моделирование, генетические алгоритмы, искусственные нейронные сети.

**Интеллектуальный анализ данных (Data Mining)**

1. Метод ближайшего соседа (NN – Nearest Neighbor)

С целью определения степени сходства на множестве параметров, используемых для описания прецедентов и текущей ситуации, вводится определенная метрика.

Основными преимуществами данного метода являются простота реализации и универсальность.

К существенным недостаткам метода можно отнести сложность выбора метрики для определения степени сходства и требовательность к ресурсам (для больших БП)

kNN – k ближайших соседей

Термин Data Mining получил свое название из двух понятий: поиска ценной информации в большом наборе информации (data) и добычи полезных ископаемых (mining).

Эти процессы требуют или просеивания огромного количества сырого материала, или разумного исследования и поиска искомых ценностей.

Термин «Data Mining» введен Григорием Пятецким-Шапиро в 1989 году.

Интеллектуальный анализ данных (Data Mining) – выявление скрытых закономерностей или взаимосвязей между переменными в больших массивах необработанных данных.

Data Mining – мультидисциплинарная область, возникшая и развивающаяся на базе таких наук как прикладная статистика, распознавание образов, искусственный интеллект, теория баз данных и др.

Иногда используют термин – Инженерия знаний (рус. Вариант)

К Data Mining относятся: теория БД, машинное обучение, алгоритмизация, искусственный интеллект, распознавание образов, визуализация, статистика, другие дисциплины.

Суть и цель технологии Data Mining можно охарактеризовать так: это технология, которая предназначена для поиска в больших объемах данных неочевидных, объективных и полезных на практике закономерностей.

**Последовательность операций KDD**

1. Консолидация данных – процесс их извлечения из различных источников (online-систем, СУБД, файлов, Интернета и т. д.) и загрузка в централизованное хранилище данных.
2. Подготовка анализируемых выборок данных (в том числе, обучающих), загрузка их из хранилища данных или других источников в аналитическое приложение.
3. Очистка данных от факторов, мешающих их корректному анализу (шум, аномальные значения, дубликаты, противоречия, пропуски, фиктивные значения и т. д.)
4. Трансформация – оптимизация данных для решения определенной задачи. Обычно на данном этапе выполняется исключение незначащих факторов, снижения размерности входных данных, нормализация, обогащение (дополнение-augmentation).
5. Анализ данных – применение методов и технологий Data Mining: построение и обучение моделей (например, нейронных сетей, деревьев решений и др.), решение задач классификации, регрессии, кластеризации, прогнозирования и т. д.
6. Интерпретация и визуализация результатов анализа, их применение в бизнес-приложениях.

**Характеристика KDD**

Knowledge Discovery in Databases не задает набор методов обработки или пригодные для анализа алгоритмы, он определяет последовательность действий, которую необходимо выполнить для того, чтобы из исходных данных получить знания. Данный подход универсальный и не зависит от предметной области.

**Data engineering**

Организация ETL/ELT pipeline, DB, Data Lake, Data Warehouse

Big Data: Apache Hadoop, Spark, Airlow

**Рассуждения на основе прецедентов**

Case-Based Reasoning / CBR

Прецедент – случай, имевший место ранее и служащий примером или оправданием для последующих случаев подобного рода.

Вывод на основе прецедентов является подходом, позволяющим решить новую задачу, используя или адаптируя решение уже известной задачи.

Преимущества:

1. Возможность напрямую использовать опыт, накопленный системой без интенсивного привлечения эксперта в той или иной предметной области.
2. Отсутствует необходимость полного и углубленного рассмотрения знаний о конкретной предметной области.
3. Возможно применение эвристик, повышающих эффективность решения задач.

Недостатки:

1. При описании прецедентов обычно ограничиваются поверхностными знаниями о предметной области.
2. Большое количество прецедентов (большой размер базы прецедентов) может привести к снижению производительности системы.
3. Выбор критериев для индексации и сравнения прецедентов является сложной задачей.
4. Невозможность получения решения задач, для которых нет прецедентов ил степень их сходства (подобия) меньше заданного порогового значения.

**База прецедентов**

Для представления прецедентов достаточно простого параметрического представления, т. е. представления прецедента в виде набора параметров с конкретными значениями и решения

CASE(X1, X2,…Xn; R)

**Основные методы извлечения прецедентов**

1. Метод ближайшего соседа (NN – Nearest Neighbor)

С целью определения степени сходства на множестве параметров, используемых для описания прецедентов и текущей ситуации, вводится определенная метрика.