

**Математические модели,
постановки задач,
алгоритмы обучения,
оценки решающих правил**

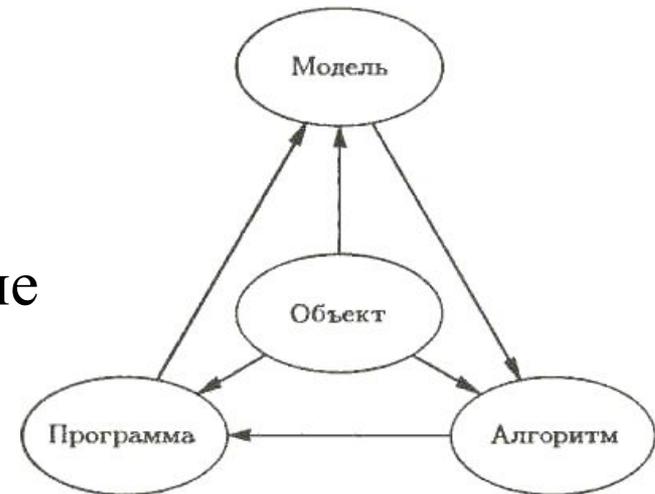




- ❖ Классификация моделей, прямая и обратная задачи, виды моделирования.
- ❖ Процесс моделирования, критерий выбора.
- ❖ Стандартные постановки основных задач индуктивного формирования баз знаний.
- ❖ Алгоритмы обучения классификации, их характеристики и способы сравнения.



Математическая модель – математическое представление реальности, один из вариантов модели, как системы, исследование которой позволяет получать информацию о некоторой другой системе.



- Замена объекта исследования его моделью
- Связь с реальностью – гипотезы, идеализация, упрощение
- Методы, как правило, описывают идеальный объект
- Универсальные модели разного уровня адекватности

Классификация моделей



Формальная классификация	Линейные или нелинейные модели Сосредоточенные или распределенные системы Детерминированные или стохастические (вероятностные) Статические или динамические Дискретные или непрерывные Однородные или неоднородные
По способу представления объекта	Структурные и функциональные
По степени формализации	Содержательные и формальные
Содержательная классификация моделей	Гипотеза, феноменологическая модель, приближение, упрощение, эвристическая модель, аналогия, мысленный эксперимент, демонстрация возможности
По наличию параметров	Модели без параметров или модели с параметрами

Прямая и обратная задачи математического моделирования



Типы задач

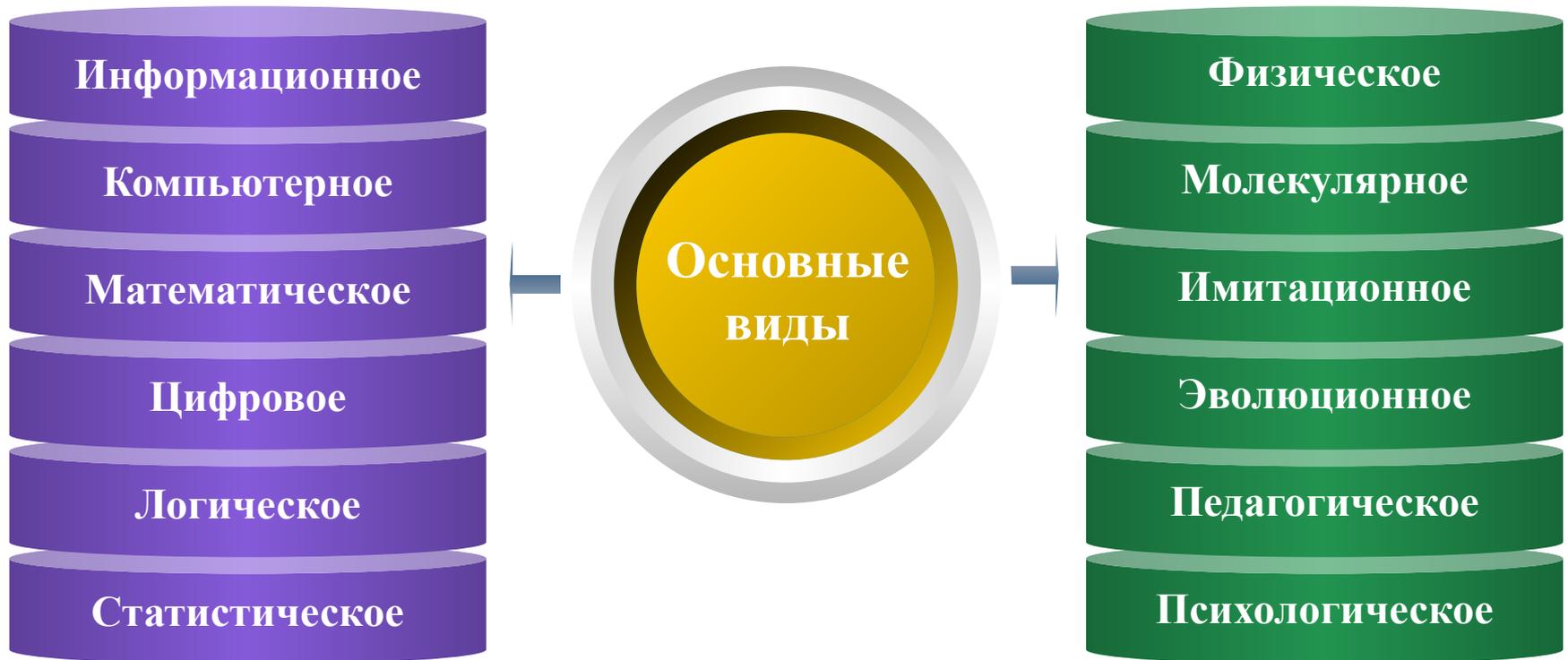
Прямая задача.

Структура модели и все ее параметры считаются известными, главная задача – провести исследование модели для извлечения полезного знания об объекте.

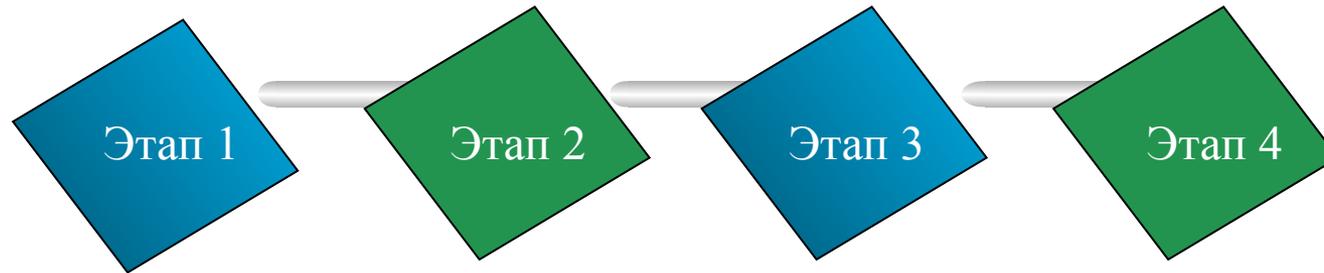
Обратная задача.

Известно множество возможных моделей и нужно выбрать конкретную модель на основании некоторых данных об объекте. Чаще всего структура модели известна, и необходимо определить некоторые неизвестные параметры.

Виды моделирования



Процесс моделирования



Сбор знаний об объекте. Построение модели конкретной сложности. Решение вопроса о сходстве оригинала и модели.

Модель как объект исследования. Модельный эксперимент. Результат этапа – получение совокупности знаний о модели.

Формирование множества знаний. Корректировка знаний с учетом не учтенных в модели свойств оригинала.

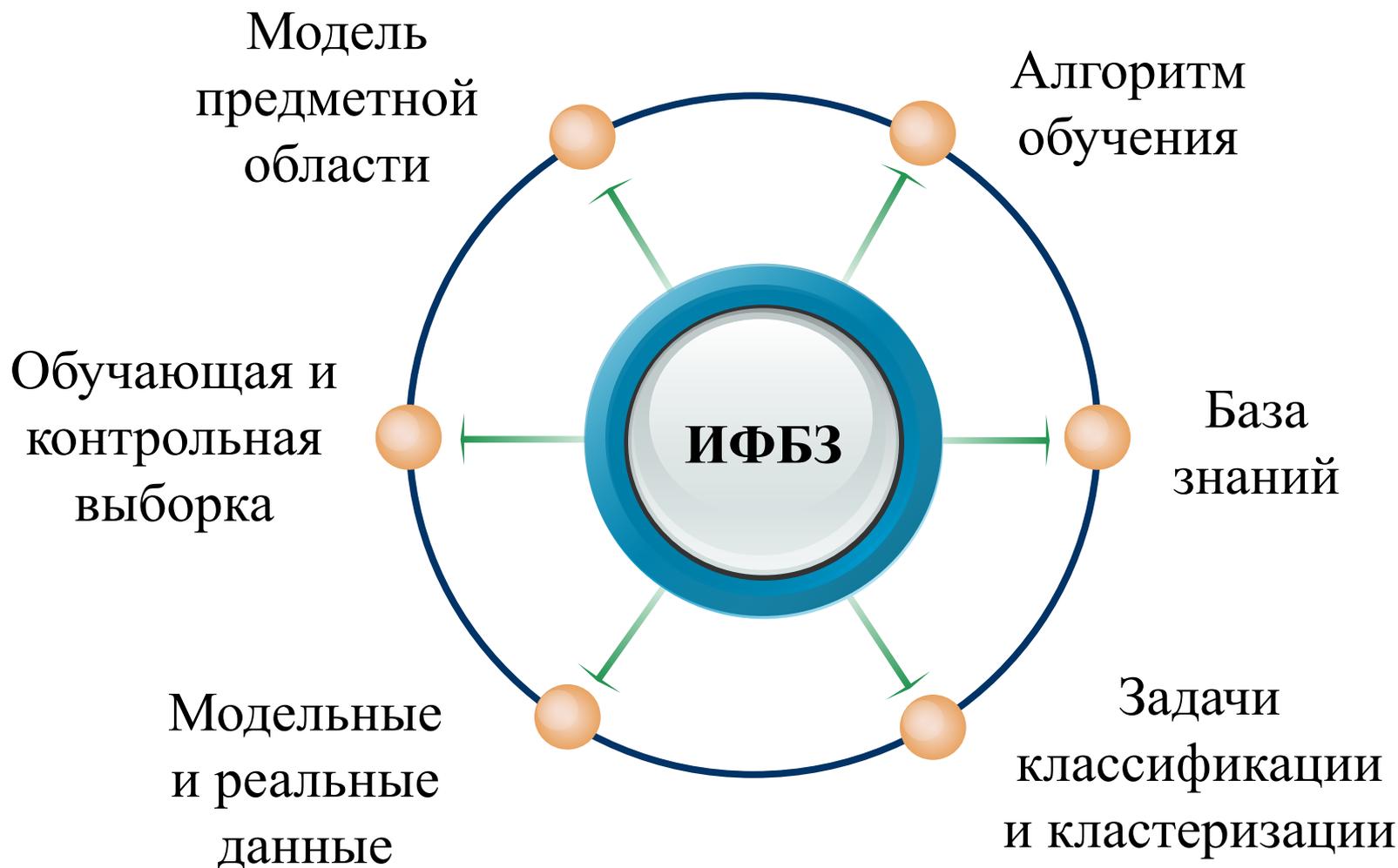
Практическая проверка полученных с помощью модели знаний. Построение обобщающей теории объекта-оригинала.



1. Внешние критерии:
адекватность, непротиворечивость, полнота, точность, универсальность
2. Внутренние критерии:
сложность (количественная), интерпретируемость (качественная)

Отбор признаков осуществляется по принципу их значимости с точки зрения удовлетворения критериям к выбору (построению) модели. Бывают случаи, когда объекты описываются временными рядами, сигналами, изображениями, видеорядами, текстами, попарными отношениями сходства или интенсивности взаимодействия.

Индуктивное формирование баз знаний (ИФБЗ)



Стандартные постановки основных задач ИФБЗ



Для некоторого множества моделей зависимости, к которому относится неизвестная зависимость между классами и объектами, **разработать алгоритм классификации**, который на основе описания объектов обучающей выборки **строит решающее правило**, вероятность правильной классификации которого любых новых объектов как можно выше.



Используя некоторую метрику, разработать алгоритм кластеризации, который на основе описания объектов обучающей выборки **разбивает обучающую выборку** на непересекающиеся подмножества, называемые кластерами, так, чтобы каждый кластер состоял из схожих объектов, а объекты разных кластеров существенно отличались, и **строит описания кластеров**, позволяющие относить к ним новые объекты.



Учитель – либо сама обучающая выборка, либо тот, кто указал на объектах обучающей выборки их правильные классы.

**Правильная
классификаци
я**

– если одним из классов объекта, выдаваемых решающим правилом, является правильный класс.

**Точная
классификаци
я**

– если решающее правило выдает для объекта единственный класс.

Отличия задач классификации и кластеризации



Задача кластеризации отличается от задачи классификации тем, что в первом случае разбиение множества объектов на классы неизвестно, и поэтому для объектов обучающей выборки правильные классы не могут быть заданы.

Задача кластеризации сводится к разбиению обучающей выборки на непересекающиеся подмножества, называемые кластерами, так, чтобы каждый кластер состоял из схожих объектов, а объекты разных кластеров существенно отличались, а также к построению описания кластеров, позволяющим относить к этим кластерам новые объекты.



Индуктивная база знаний

формирование



Модель



**Постановка
задачи**



**Алгоритм
обучения**



**Обучающая
выборка**



Алгоритм классификации (алгоритм обучения классификации) – это **отображение, которое по обучающей выборке строит решающее правило.** Нужно построить такой алгоритм, вероятность правильной классификации которого новых объектов имеет возможно большее значение.

Поскольку задача поиска наибольшего значения для всех возможных алгоритмов классификации и обучающих выборок не имеет шансов быть решенной, как правило, рассматривается более конкретная постановка задачи классификации, например, задача чемпионата мира среди алгоритмов классификации.

Задача чемпионата мира среди алгоритмов классификации (слабая)



Пусть имеются алгоритмы классификации L_{m_1}, \dots, L_{m_z} , для моделей m_1, \dots, m_z соответственно, и обучающая выборка t . Для модели m требуется построить такой алгоритм классификации L_m , что для заданной обучающей выборки t имеет место $P(L_m(t)) > P(L_{m_1}(t)), \dots, P(L_m(t)) > P(L_{m_z}(t))$.

В соответствии с этой постановкой задачи проводятся чемпионаты мира среди алгоритмов классификации (L_m считается победителем среди участников L_{m_1}, \dots, L_{m_z}).

Во многих работах предлагаемый алгоритм сравнивается таким способом с другими известными алгоритмами классификации.

Задача чемпионата мира среди алгоритмов классификации (сильная)



Если в постановке задачи рассматривается множество обучающих выборок $\mathbf{t}_1, \dots, \mathbf{t}_v$, то такую постановку можно назвать «сильной». В последнем случае алгоритм классификации L_m считается победителем среди участников L_{m_1}, \dots, L_{m_z} , если его вероятность правильной классификации превосходит вероятности других алгоритмов на всех обучающих выборках $\mathbf{t}_1, \dots, \mathbf{t}_v$.

Сильная постановка задачи классификации поддерживается распределенной системой тестирования алгоритмов классификации «Полигон алгоритмов».