

Лекция 1. Вводная. Обзор состояния области.
Проблемы машинного обучения и индуктивного вывода знаний.

1. Не известно, что за информация содержится в данных и как ее извлекать. Для физических величин вводилось понятие размерности, 5 литров или 5 метров или 5 баллов говорят об интерпретации числового значения и о том, что с ним можно делать. Для нефизических величин такой фиксированной интерпретации нет. Например, температура тела человека или почвы или курицы в духовке – говорят о совершенно разных величинах – они являются косвенными измерениями некоторых скрытых и неизвестных нам величин, которые интерпретируются в терминах предметной области. Как узнать, что можно делать с величиной? (1) найти интерпретацию величины в онтологии предметной области; (2) обратиться к теории измерений.
 - a. Искажение информации существующими методами;
 - b. Надо учитывать тип шкал величин и допустимые преобразования шкал;
 - c. Не инвариантность методов и не интерпретируемость их результатов;
 - d. Для определения извлекаемой информации и шкал величин надо обратиться к теории измерений и к онтологии предметной области;
 - e. Для корректной обработки надо извлечь из данных интерпретируемую информацию и представить ее в символьном виде;
 - f. Базы данных не в состоянии поддерживать извлечение информации в логике первого порядка;
 - g. Методы машинного обучения обнаруживают только заранее заданные классы гипотез и моделей. Нет возможности исследовать и познавать предметную область.
 - h. Задача: разработать инвариантный и достаточно общий метод в логике первого порядка.
2. Не известно, что такое закон природы.
 - a. Определение законов в Теории Измерений и Теории Физических структур;
 - b. Обнаруживать законы надо с одновременным формированием понятий.
3. Проблема логического вывода знаний.
 - a. Индуктивный вывод знаний и теории предметной области;
 - b. Экспертные системы, проблемы их создания:
 - i. неизвлекаемость интуитивных знаний;
 - ii. проблема извлечения знаний из эксперта – проблема огромного числа вопросов (одна методика будет рассказана);
 - iii. нет адекватного вывода предсказаний и прогнозов из знаний;
 - c. Проблема формализации предсказания;
 - d. Статистическая двусмысленность индуктивно выводимых знаний;
 - e. Синтез логики и вероятности;
 - f. Семантический вероятностный вывод (СВВ) и решение проблем:
 - i. индуктивного вывода теории ПО;
 - ii. вывода предсказаний;
 - iii. решение проблемы статистической двусмысленности;
4. Реляционный подход к извлечению знаний из данных и компьютерное познание.
 - a. Система Discovery реализует семантический вероятностный вывод;
 - b. Разработан реляционный подход к методам извлечения знаний, обладающий следующими свойствами:
 - i. использует всю необходимую информацию в данных;
 - ii. инвариантен;
 - iii. обнаруживает: теорию, максимально вероятные знания, максимально специфические знания, вероятностную аппроксимацию теории – в точности все необходимые знания;
 - iv. аппроксимирует другие методы.
 - c. Решены интересные прикладные задачи;
 - d. Проект Microsoft объединения методов машинного обучения.
5. Формализация когнитивных процессов и моделей нейрона и работы мозга на основе СВВ.
6. Пересмотр искусственного интеллекта.