

Анализ научных данных: от концепций к результатам

АНДРЕЙ ВЛАДИМИРОВИЧ АЛЕКСЕЕНКО

КАНДИДАТ БИОЛОГИЧЕСКИХ НАУК, ДОЦЕНТ (KAROLINSKA INSTITUTET, STOCKHOLM), ПО ОБРАЗОВАНИЮ БИОЛОГ СО СПЕЦИАЛИЗАЦИЕЙ НА ГЕНЕТИКЕ РАСТЕНИЙ. ПОСЛЕДНИЕ 20 ЛЕТ ЗАНИМАЕТСЯ КОЛИЧЕСТВЕННЫМ АНАЛИЗОМ БИМЕДИЦИНСКИХ ДАННЫХ. АВТОР ОКОЛО 50 НАУЧНЫХ РАБОТ. СОЗДАТЕЛЬ МЕТОДОВ АНАЛИЗА МНОЖЕСТВ И ГРАФОВ, А ТАКЖЕ ВЕБ-САЙТОВ И ДРУГИХ ПУБЛИЧНО ДОСТУПНЫХ РЕСУРСОВ. ПРЕПОДАВАТЕЛЬ И ОРГАНИЗАТОР РЯДА КУРСОВ ПО ПРИКЛАДНОЙ СТАТИСТИКЕ И СИСТЕМНОЙ БИОЛОГИИ.

Современная методология анализа данных может предстать начинающему исследователю оторванной от изученных стандартных курсов математики и статистики. Цель курса – установление связей между теоретическими понятиями и практическими целями и методами исследований в разных областях, от биомедицины до общественных наук. Обзор ряда методов и общение преподавателя со студентами поможет уточнить интересы, задачи курса и терминологию – как междисциплинарно, так и между русским и английским доменами.

Курс

1. История анализа данных. Концепции и подходы*. Дискуссия: содержание курса, обмен опытом и мнениями.
2. Наблюдения. Переменные и параметры. Информация, неопределенность и их соотношение. Форматы представления данных. Обзор принципов и методов одно- и многомерного анализа.
3. Множества. Enrichment: статистически значимые отклонения от пересечения множеств.
4. Семинар по теме (3) и практическое занятие с использованием онлайн-ресурсов.

5. Графы и сети (очень большие графы) как подход к визуализации и анализу данных. Основные понятия, типы и структурные элементы графов. Алгоритмы анализа, их назначение и вычислительная сложность. Статистическая значимость результатов. Информационное содержание и emerging properties.
 6. Семинар по теме (3) и практическое занятие с использованием онлайн-ресурсов.
 7. Временные ряды. Причинно-следственные связи.
 8. Достоверность. Воспроизводимость. Робастность. Валидация. Мета-анализ. Модели, их назначение и создание, оценка качества и практическое применение. Over-fitting и dimensionality curse. Сравнительный обзор моделирования на различных уровнях.
 9. Семинар по теме (8).
 10. Свободная дискуссия. Вопросы-ответы.
 11. Итоговый экзамен: письменная работа.
- * формат: лекция, если не указано иное.

Пререквизиты к слушателям

Знание английского языка на уровне уверенного чтения. Опыт работы с данными или намерение использовать анализ данных для конкретных задач. Умение программировать или научиться этому в будущем не требуются, но приветствуется.

Список литературы

Литература по курсу будет даваться в основном на английском языке, в виде отдельных статей или небольших выдержек.