



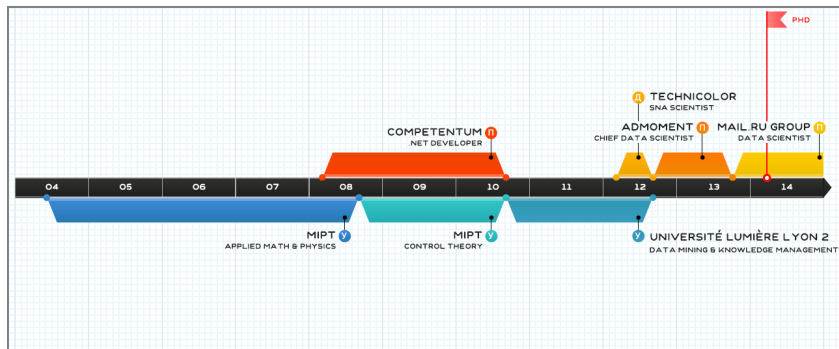
ТЕХНОСФЕРА

Лекция 1 Задачи Data Mining

Николай Анохин

23 сентября 2014 г.

Николай Анохин



e-mail: n.anokhin@corp.mail.ru

тел.: +7 (903) 111-44-60

План лекции

Структура курса

Что такое Data Mining

Структура курса

Модуль 1

1. Задачи Data Mining (Николай Анохин)
2. Задача кластеризации и ЕМ-алгоритм (Николай Анохин)
3. Различные алгоритмы кластеризации (Николай Анохин)^H
4. Задача классификации (Николай Анохин)
5. Naïve Bayes (Николай Анохин)
6. Линейные модели (Николай Анохин)
7. Метод опорных векторов (Николай Анохин)^{HP}

Модуль 2

1. Снижение размерности пространства (Владимир Гулин)
2. Алгоритмические композиции 1 (Владимир Гулин)
3. Алгоритмические композиции 2 (Владимир Гулин)^H
4. Нейросети, обучение с учителем (Павел Нестеров)^H
5. Нейросети, обучение без учителя (Павел Нестеров)
6. Нейросети, глубокие сети (Павел Нестеров)

Контроль знаний

ДЗ

4 домашних задания на 6-8 часов самостоятельной работы каждое
(4x15 баллов)

Экзамен

Презентация и защита семестрового проекта
(40 баллов)

Правила

- + Можно задавать вопросы по ходу лекции
- + Можно входить и выходить, не мешая коллегам
- Нельзя нарушать порядок в аудитории
- Нельзя разговаривать по телефону
 - ▶ Общение с преподавателем на “Вы”

Ваши правила?

DM как KDD

Data Mining

Процесс извлечения знаний из различных источников данных, таких как базы данных, текст, картинки, видео и т.д. Полученные знания должны быть *достоверными, полезными и интерпретируемыми*.

DM как моделирование

Data Mining

Процесс построения модели, хорошо описывающей закономерности, которые порождают данные.

Подходы к построению моделей

- ✓ статистический
- ✓ на основании машинного обучения
- ✗ вычислительный

DM и DS

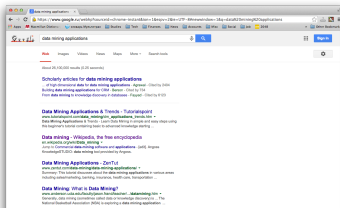
Data Scientist

Person who is better at statistics than any software engineer and better at software engineering than any statistician
(J. Wills, Data Scientist at Cloudera Inc.)

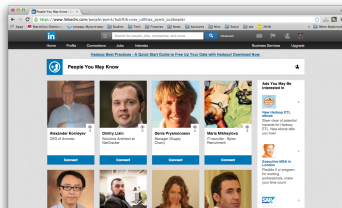
Data-

- ✗ -architecture
- ✗ -acquisition
- ✓ -analysis
- ✗ -archiving

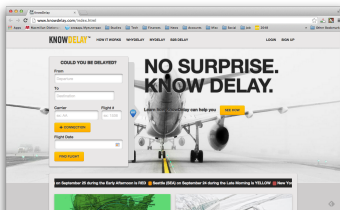
Success stories



(a) Google



(b) LinkedIn



(c) KnowDelay



(d) Handwritten digits

Вопросы

