MAKE Decisions
Medical Information Science for Decision Support

Assoc. Prof. Dr. Andreas HOLZINGER (Medical University Graz)

https://hci-kdd.org/mini-course-make-decisions-practice

Day 1 > Part 1 > 19.09.2018

Information Sciences meets Life Sciences
Andreas Holzinger: Background

- PhD in Cognitive Science 1998
- Habilitation Computer Science 2003
- Lead Holzinger Group HCI-KDD
  www.hci-kdd.org
- Visiting Professor for Machine Learning in Health Informatics: TU Vienna, Univ. Verona, UCL London, RWTH Aachen
- Research Statement see:
  Brain Informatics, 3, 1-13,
  doi:10.1007/s40708-016-0042-6
- Most recent:
  Informatik-Spektrum,
At the end of this course you will ...

... be fascinated to see our world in **data sets**;

... understand the differences between **data, information and knowledge**

... be aware of some problems and challenges in **biomedical informatics**

... understand the importance of the concept of **probabilistic information** \( p(x) \)

... know what **AI/Machine Learning** can (not) do

... have some fundamental insight into medical information science for **decision making**
Overview

Primer on Probability & Information

Day 1 - Fundamentals

01 Information Sciences meets Life Sciences

02 Data, Information and Knowledge

03 Decision Making and Decision Support

04 From Expert Systems to Explainable AI

Day 2 – Hot Topics

05 Methods of Explainable-AI

Groupwork: Planning of a 500 bed Hospital - Bringing AI into the workflows

Plenary: Presenting the developed concepts
Agenda

- 01 What is the HCI-KDD approach?
- 02 Application Area: Health Informatics
- 03 Probabilistic Information
- 04 Automatic Machine Learning
- 05 Interactive Machine Learning
- 06 Key Problems in Biomedical Informatics
01 What is the **HCI-KDD** approach?
ML is a very practical field – algorithm development is at the core – however, successful ML needs a concerted effort of various topics ...
MAchine Learning & Knowledge Extraction MAKE

(Safety) 4 - Privacy, Data Protection, Safety & Security

3 - Visualization

2 - Learning

1 - Data

(Space and Time) 5 - Network, 6-Topology, 7-Entropy

Cognitive Science AND Computer Science

- Cognitive Science → human intelligence
- Computer Science → computational intelligence
- Human-Computer Interaction → the bridge

To reach a level of usable intelligence we need to ...

- 1) learn from prior data
- 2) extract knowledge
- 3) generalize, i.e. guessing where a probability mass function concentrates
- 4) fight the curse of dimensionality
- 5) disentangle underlying explanatory factors of data, i.e.
- 6) understand the data in the context of an application domain
“Solve intelligence – then solve everything else”

Demis Hassabis, 22 May 2015
The Royal Society,
Future Directions of Machine Learning Part 2

https://youtu.be/XAbLn66iHcQ?t=1h28m54s
How far are we already?

Compare your best ML algorithm with a seven year old child ...

Not our Goal: Humanoid AI

Humanoid AI ≠ Human-level AI
02 Application Area
Health Informatics
Why is this application area complex?
In medicine we have two different worlds ...

Our central hypothesis:
Information may bridge this gap

Where is the problem in building this bridge
Heterogeneity

Dimensionality

Complexity

Uncertainty

03 Probabilistic Information $p(x)$
Probability theory is nothing but common sense reduced to calculation...

Pierre Simon de Laplace (1749-1827)
Newton, Leibniz, ... developed calculus – mathematical language for describing and dealing with rates of change

Bayes, Laplace, ... developed probability theory - the mathematical language for describing and dealing with uncertainty

Gauss generalized those ideas
Learning representations \((\theta, h)\) from observed data

**Observed data:**

\[ \mathcal{D} = x_1:n = \{x_1, x_2, \ldots, x_n\} \]

**Feature Parameter:**

\[ \theta \quad \text{or hypothesis } h \quad h \in \mathcal{H} \]

**Prior belief \(\approx\) prior probability of hypothesis \(h\):**

\[ p(\theta), p(h) \]

**Likelihood \(\approx\) \(p(x)\) of the data that \(h\) is true**

\[ p(\mathcal{D}|\theta), p(d|h) \]

**Data evidence \(\approx\) marginal \(p(x)\) that \(h = \text{true}\)**

\[ p(\mathcal{D}) \sum_{h \in \mathcal{H}} p(d|h) \cdot p(h) \]

**Posterior \(\approx\) \(p(x)\) of \(h\) after seen (“learn”) data \(d\)**

\[ p(\theta|\mathcal{D}), p(h|d) \]

\[
posterior = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}} \quad p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta) \cdot p(\theta)}{p(\mathcal{D})}
\]

\[
p(h|d) = \frac{p(d|h) \cdot p(h)}{\sum_{h \in \mathcal{H}} p(d|h) \cdot p(h)}
\]
Learning and Probabilistic Inference (Prediction)

\( d \) ... data
\( \mathcal{H} \) ... \{\( H_1, H_2, ..., H_n \}\)
\( h \) ... hypotheses

\[
p(h|d) = \frac{p(d|h) \cdot p(h)}{\sum_{h \in \mathcal{H}} p(d|h) \cdot p(h)}
\]

Prior Probability
Likelihood
Posterior Probability

Problem in \( \mathbb{R}^n \rightarrow \text{complex} \)
Why is this relevant for health informatics?
Reasoning under uncertainty: Decision Making

- Take patient information, e.g., observations, symptoms, test results, -omics data, etc. etc.
- Reach conclusions, and **predict** into the future, e.g. how likely will the patient be ...
- Prior = belief before making a particular observation
- Posterior = belief after making the observation and is the prior for the next observation – intrinsically incremental

\[
p(x_i|y_j) = \frac{p(y_j|x_i)p(x_i)}{\sum p(x_i, y_j)p(x_i)}
\]
GP = distribution, observations occur in a cont. domain, e.g. t or space

Bayesian Optimization 2
Fully automatic $\rightarrow$ Goal: Taking the human out of the loop

Algorithm 1 Bayesian optimization

1: $\text{for } n = 1, 2, \ldots \text{ do}$
2: $\text{select new } x_{n+1} \text{ by optimizing acquisition function } \alpha$
   
   $x_{n+1} = \arg \max_x \alpha(x; \mathcal{D}_n)$
3: $\text{query objective function to obtain } y_{n+1}$
4: $\text{augment data } \mathcal{D}_{n+1} = \{ \mathcal{D}_n, (x_{n+1}, y_{n+1}) \}$
5: $\text{update statistical model}$
6: $\text{end for}$

Taking the human out of the loop: A review of Bayesian optimization. 
04 aML
Example for aML: Recommender Systems
Fully automatic autonomous vehicles (Google car)

... and thousands of industrial aML applications ...

Big Data is necessary for aML!

\[ x^* = \arg \min_x f(x; W, H), \text{ subject to } \|x\|_2 = 1. \]


Limitations of Deep Learning approaches

- **Computational resource intensive** (supercomps, cloud CPUs, **federated learning**, ...)
- **Black-Box approaches** – lack **transparency**, do not foster trust and acceptance among end-user, **legal aspects** make “black box” difficult!
- **Non-convex**: difficult to set up, to train, to optimize, needs a lot of expertise, error prone
- Very bad in dealing with **uncertainty**
- **Data intensive**, needs often millions of training samples ...
When does aML fail ...

- Sometimes we do not have “big data”, where aML-algorithms benefit.
- Sometimes we have
  - Small amount of data sets
  - Rare Events – no training samples
  - NP-hard problems, e.g.
    - Subspace Clustering,
    - k-Anonymization,
    - Protein-Folding, ...

Sometimes we (still) need a human-in-the-loop
Definition of iML (Holzinger – 2016)

- iML := algorithms which interact with agents*) and can optimize their learning behaviour through this interaction

*) where the agents can be human

Sometimes we need a doctor-in-the-loop
A group of experts-in-the-loop
A crowd of people-in-the-loop
aML: taking the human-out-of-the-loop

A) Unsupervised ML: Algorithm is applied on the raw data and learns fully automatic – Human can check results at the end of the ML-pipeline

B) Supervised ML: Humans are providing the labels for the training data and/or select features to feed the algorithm to learn – the more samples the better – Human can check results at the end of the ML-pipeline

C) Semi-Supervised Machine Learning: A mixture of A and B – mixing labeled and unlabeled data so that the algorithm can find labels according to a similarity measure to one of the given groups
D) Interactive Machine Learning: Human is seen as an agent involved in the actual learning phase, step-by-step influencing measures such as distance, cost functions ...

**Constraints** of humans: Robustness, subjectivity, transfer?

**Open Questions:** Evaluation, replicability, ...

Three examples for the usefulness of the iML approach

- **Example 1: Subspace Clustering**
- **Example 2: k-Anonymization**
- **Example 3: Protein Design**


06 Key Problems in health informatics
Key Problems

- **Zillions** of different biological species (humans, animals, bacteria, virus, plants, ...);
- Enormous **complexity** of the medical domain [1];
- **Complex**, heterogeneous, high-dimensional, big data in the life sciences [2];
- Limited **time**, e.g. a medical doctor in a public hospital has only 5 min. to make a decision [3];
- Limited **computational power** in comparison to the complexity of life (and the natural limitations of the Von-Neumann architecture, ...);

What is the challenge?
Two thematic mainstreams in dealing with data ...

Time

- e.g. Entropy

- Dali, S. (1931) The persistence of memory

Space

- e.g. Topology

- Bagula & Bourke (2012) Klein-Bottle
Slide 1-4: First yeast protein-protein interaction network

Nodes = proteins
Links = physical interactions (bindings)
Red Nodes = lethal
Green Nodes = non-lethal
Orange = slow growth
Yellow = not known

First human protein-protein interaction network

Light blue = known proteins
Orange = disease proteins
Yellow ones = not known yet

Conclusion: Five decades of Health Informatics

- 1960+ Medical Informatics (Early “AI”)
  - Focus on data acquisition, storage, accounting (typ. “EDV”), Expert Systems
  - The term was first used in 1968 and the first course was set up 1978!

- 1985+ Health Telematics (AI winter)
  - Health care networks, Telemedicine, CPOE-Systems, ...

- 1995+ Web Era (AI is “forgotten”)
  - Web based applications, Services, EPR, distributed systems, ...

- 2005+ Success statistical learning (AI renaissance)
  - Pervasive, ubiquitous Computing, Internet of things, ...

- 2010+ Data Era – Big Data (super for AI)
  - Massive increase of data – data integration, mapping, ...

- 2020+ Information Era – (towards explainable AI)
  - Sensemaking, disentangling the underlying concepts, causality, ...
Thank you!
Exam Questions
Appendix
Learning and Probabilistic Inference (Prediction)

\[ p(h | d) = \frac{p(d | h) \cdot p(h)}{\sum_{h \in H} p(d | h) \cdot p(h)} \]

- **\( d \)**: data
- **\( h \)**: hypotheses
- **\( \mathcal{H} \)**: \{\( H_1, H_2, ..., H_n \}\)
- \( \forall h, d \)

Prior Probability

Likelihood

Posterior Probability

Problem in \( \mathbb{R}^n \to \) complex

- \( P(h) \): prior belief (probability of hypothesis \( h \) before seeing any data)
- \( P(d | h) \): likelihood (probability of the data if the hypothesis \( h \) is true)
- \( P(d) = \sum_h P(d | h)P(h) \): data evidence (marginal probability of the data)
- \( P(h | d) \): posterior (probability of hypothesis \( h \) after having seen the data \( d \))
What is biomedical informatics?

- **Biomedical informatics (BMI)** is the interdisciplinary field that studies and pursues the effective use of biomedical data, information, and knowledge for scientific problem solving, and decision making, motivated by efforts to improve human health

Computational Sciences meet Life Sciences

http://www.bioinformaticslaboratory.nl/twiki/bin/view/BioLab/EducationMIK1-2
What is a computer?
Computer: Von-Neumann Architecture

- External Memory
  - Long term: HDD, CD, Stick etc.

- Internal Memory
  - Short term: RAM
  - Long term: ROM

- Controller
  - (BIOS, OS, AP)

- CPU

- Monitor
  - Printer
  - Modem
  - Network etc.

- Input
  - Keyboard
  - Mouse
  - Graphic Pad
  - Microphone
  - Modem
  - Network etc.

- Output

Holzinger (2002), 90-134

Digital Power :=

\[
\text{communication} \times \text{computing} \times \text{storage} \times \text{content}
\]

- „fiber law“ doubles every 9 months
- „Moore's law“ doubles every 18 months
- „disk law“ doubles every 12 months
- „community law“ \(2^n\) wherein \(n\) is the \# of people

Computer cost/size versus Performance

Beyond Moore’s Law -> biological computing

... using technology to augment human capabilities for structuring, retrieving and managing information

Example Pervasive Computing in the Hospital

Smart Objects in the pathology

Holzinger et al. (2005)
The medical world is mobile (Mocomed)

1970 Turning Knowledge into Data

Photo by Institute of Medical Informatics, Graz (1970)
Overview

Primer on Probability & Information

Day 1 - Fundamentals

01 Information Sciences meets Life Sciences

02 Data, Information and Knowledge

03 Decision Making and Decision Support

04 From Expert Systems to Explainable AI
The Travelling Snakesman
NEW Version v 1.1 - 18.4.2018