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

Mini-Course Syllabus

- 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

Reading on Paper or on any electronic device

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 ...
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
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

Health is a complex area

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

Main problems ...

Heterogeneity

Dimensionality

Complexity

Uncertainty

03 Probabilistic Information \( p(x) \)


Probability theory is nothing but common sense reduced to calculation...

---

**Learning representations** \((\theta, h)\) from observed data

- **Observed data:** \(\mathcal{D} = x_{1:n} = \{x_1, x_2, \ldots, x_n\}\)
- **Feature Parameter:** \(\theta\) or hypothesis \(h\) \(h \in \mathcal{H}\)
- **Prior belief** \(\approx\) prior probability of hypothesis \(h\):
  \[ p(\theta) \quad p(h) \]
- **Likelihood** \(\approx p(x)\) of the data that \(h\) is true:
  \[ p(\mathcal{D} | \theta) \quad p(d | h) \]
- **Data evidence** \(\approx\) marginal \(p(x)\) that \(h = \) true
  \[ p(\mathcal{D}) \quad \sum_{h \in \mathcal{H}} p(d | h) \times p(h) \]
- **Posterior** \(\approx p(x)\) of \(h\) after seen (“learn”) data \(d\)
  \[ p(\theta | \mathcal{D}) \quad p(h | d) \]

**Posterior**

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

---

**Learning and Probabilistic Inference (Prediction)**

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

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

- **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

---

---
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

\[ p(x_i|y_j) = \frac{p(y_j|x_i)p(x_i)}{\sum p(x_i, y_j)p(x_i)} \]

Learning from data


Scaling to high-dimensions is the holy grail in ML


Fully automatic → Goal: Taking the human out of the loop

Algorithm 1: Bayesian optimization
1: for \( n = 1, 2, \ldots \) do
2: select new \( x_{n+1} \) by minimizing acquisition function \( \alpha \)
   \[
   x_{n+1} = \arg \max_{x \in \mathbb{R}} \alpha(x; \mathcal{D}_n)
   \]
3: query objective function to obtain \( f_{n+1} \)
4: augment data \( \mathcal{D}_{n+1} = \{ \mathcal{D}_n, (x_{n+1}, f_{n+1}) \} \)
5: update statistical model
6: end for

04 aML

Fully automatic autonomous vehicles (Google car)


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

Big Data is necessary for aML!

10 million 200 x 200 px images downloaded from Web

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


Deep Convolutional Neural Network Pipeline


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, ...

Consequently ...

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, ...


06 Key Problems in health informatics

- **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, ...);

**Our World in Data - Macroscopic Structures**

What is the challenge?

ESO, Atacama, Chile (2011)

**Our World in Data – Microscopic Structures**

**Knowledge Discovery from Data**

**Time**

e.g. Entropy

Dali, S. (1931) The persistence of memory

**Space**

e.g. Topology

Bagula & Bourke (2012) Klein-Bottle

**Two thematic mainstreams in dealing with data ...**


Non-Natural Network Example: Blogosphere


Social Behavior Contagion Network

Human Disease Network -> Network Medicine


Identifying Networks in Disease Research


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

Learning and Probabilistic Inference (Prediction)

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

- **Prior Probability**: \( p(h) \)
- **Likelihood**: \( p(d|h) \)
- **Posterior Probability**: \( p(h|d) \)

Problem in \( \mathbb{R}^{71} \) → complex

- **Prior belief**: \( p(h) \)
- **Likelihood**: \( p(d|h) \)
- **Data evidence**: \( p(d) = \sum_{h} p(d|h) \cdot p(h) \)
- **Posterior**: \( p(h|d) \)

Appendix

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.

Computer cost/size versus Performance

- Vast reduction in cost, but enormous capability increase.


Beyond Moore’s Law -> biological computing

- Memory: $10^7$ bit
- Logic: $>10^6$ bit
- Power: $10^{-13}$ W
- Heat: $10^{-8}$ W/cm²
- Energy/task: $10^{-10}$ J
- Task time: 2400 s = 40 min

Equivalent to $10^{11}$ output bits


From mainframe to Ubiquitous Computing

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


Ubiquitous Computing – Smart Objects

Slide 1-34 Example: Pervasive Health Computing


Health Informatics

Ambient Assisted Living - pHealth


Health Informatics

Example Pervasive Computing in the Hospital


Health Informatics

Smart Objects in the pathology

Health Informatics
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 medical world is mobile (Mocomed)


1970 Turning Knowledge into Data

Photo by Institute of Medical Informatics, Graz (1970)

Health Informatics

Health Informatics