Welcome to the Holzinger Group HCI-KDD
Part 1: What is the HCI-KDD approach and what is aML vs. iML?

a.holzinger@hci-kdd.org
http://hci-kdd.org/scientific-working-for-students
The “best” is the enemy of the “good” – whenever you try to be “perfect” – there is the danger that you finalize nothing*) ...”

*) zero, nada, null

François-Marie Arouet (1694 – 1778) known as “Voltaire”
Science is to test crazy ideas – Engineering is put these ideas into Business!
At the end of the whole seminar you should
- be aware of the HCI-KDD approach
- know the interests of the Holzinger group
- have an overview on possible research topics
- be familiar with the formal PhD requirements
- understand how to carry out scientific research
- know how to write scientific papers
- most of all: getting started with your work
01 What is the HCI-KDD approach?
02 Application Area: Health
03 Probabilistic Information
04 Gaussian Processes
05 Automatic Machine Learning (aML)
06 Interactive Machine Learning (iML)
07 Conclusion
01 What is the HCI-KDD approach?
Machine Learning is a very practical field – at the core is algorithm development – however, successful machine learning needs a concerted effort of various fields ...
The heart in our Knowledge Discovery pipeline is ML ...

Interactive | Data Mining | Knowledge Discovery
---|---|---
6 Data Visualization | 2 Learning Algorithms | 1 Data Mapping
| 3 Graph-based Data Mining | Preprocessing | Data Fusion
| 4 Topological Data Mining | GDM | EDM
| Entropy-based Data Mining |

Privacy, Data Protection, Safety and Security

... but successful ML needs a concerted effort!

combined effort

international

without boundaries ...

Holzinger Group, HCI-KDD.org

Student Seminar, Winter 2016
Grand challenge: Transfer results from $\mathbb{R}^n$ to $\mathbb{R}^2$

Human intelligence (Cognitive Science)

Machine intelligence (Computer Science)

Holzinger, A. (2013). Human–Computer Interaction & Knowledge Discovery (HCI-KDD): What is the benefit of bringing those two fields to work together? In: Lecture Notes in Computer Science 8127 (pp. 319-328)
Cognitive Science vs. Computer Science

- **Cognitive Science → human intelligence**
  - Study the principles of *human learning*
    to understand biological intelligence

- **Human-Computer Interaction → the bridge**
  - Interacting with algorithms that learn shall enhance user friendliness and let concentrate on problem solving - Opening the “black-box” to a “glass-box”

- **Computer Science → computational intelligence**
  - Study the principles of *machine learning*
    to understand artificial intelligence
CS aims to reverse engineer human intelligence;

ML provides powerful sources of insight into *how machine intelligence* can be possible.

CS therefore raises challenges for, and draws inspiration from ML;

Insights about the human mind may help inspire new directions for ML ...
“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
What makes a machine intelligent? Cross-cutting issues

- To hear, to see, to talk, (to smell, taste, touch)
  - Speech recognition, computer vision, natural language processing (olfactory, gustatory sensors)

- To store, to represent, to access
  - Knowledge representation, semantic networks, ontologies, information retrieval

- To reason, to understand, to reflect
  - Logic, Bayesian inference, contextual understanding

- To learn from data
  - Improve with experience from previous events
Now, compare your best Machine Learning algorithm with a seven year old child ...

Learning complex concepts from a few examples

“An ultra-intelligent machine could design even better machines; there would then unquestionably be an "intelligence explosion*" and the intelligence of man would be left far behind ...

It is curious that this point is made so seldom ... outside of science fiction.”

Irving John Good, Trinity College, Oxford, 1965
Colleague of Alan Turing in Bletchley Park


*) https://intelligence.org/ie-faq/
Today ML is enormously progressing ...

- Progress is driven by the explosion in the availability of big data and low-cost computation.
- Health is amongst the biggest challenges.

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 ...
03 Probabilistic Information $p(x)$
The foundation for machine learning was laid in 1763...


\[ p(x_i) = \sum P(x_i, y_j) \]

**Thomas Bayes**

1701 - 1761

\[ p(x_i, y_j) = p(y_j|x_i)P(x_i) \]

Bayes’ Rule is a corollary of the Sum Rule and Product Rule:

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

$p(h|d) = \frac{p(d|h) \cdot p(h)}{\sum_{h' \in H} p(d|h') \cdot p(h')}$

$d$ ... data  \hspace{1cm} $H$ ... \{H_1, H_2, ..., H_n\}  \hspace{1cm} \forall h, d$ ...

$h$ ... hypothesis
Health Example
Learning from previous examples ...

The more examples we have the better ...

Bayesian Learning from data

\[
\mathcal{D} = x_{1:n} = \{x_1, x_2, \ldots, x_n\}
\]

\[
p(D|\theta)
\]

\[
p(\theta|D) = \frac{p(D|\theta) \cdot p(\theta)}{p(D)}
\]

**posterior** = \(\frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}}\)

The inverse probability allows to learn from data, infer unknowns, and make predictions.
Let the data do the work!

- Machine Learning is the development of algorithms which can learn from data
- assessment of uncertainty, making predictions
- Automating automation - getting computers to program themselves – let the data do the work!
- 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 integrated those theories
04 Gaussian Processes
From Bayesian Optimization to Gaussian Process (GP) approximation

GP = distribution, observations occur in a cont. domain, e.g. t or space

GP posterior
\[ p(f(x)|D) \propto p(D|f(x)) p(f(x)) \]

Likelihood

GP prior

Demo on how Bayesian Optimization works ...

Bayesian Optimization 3
Bayesian Optimization 5

The diagram illustrates the concept of Bayesian Optimization, which is a sequential decision-making problem where the aim is to find the maximum of a function $f(x)$ using as few evaluations as possible. The expected improvement (EI) acquisition function is shown below the function, guiding the selection of the next point to evaluate. The solid line represents the true function $f(x)$, while the dotted line shows the predicted function. The EI curve indicates the expected improvement at each point, suggesting the next informative point to test. The optimization process iteratively selects the next point to evaluate based on the highest expected improvement.
Why is this relevant for health informatics?
Reasoning under uncertainty

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

\[
p(x_i | y_j) = \frac{p(y_j | x_i) p(x_i)}{\sum p(x_i, y_j) p(x_i)}
\]
... big data is good for automatic Machine Learning
and the grand goal of aML is ...
05 aML
Today most ML-applications are using automatic Machine Learning (aML) approaches

automatic Machine Learning (aML) := algorithms which interact with agents and can optimize their learning behaviour through this interaction

Best practice examples of aML
Fully automatic autonomous vehicles ("Google car")

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

Cyber-Physical Systems (CPS):
*Tight integration of networked computation with physical systems*

- Automotive
  - E-Corner, Siemens
- Building Systems
- Telecommunications
- Transportation (Air traffic control at SFO)
- Avionics
- Instrumentation (Soleil Synchrotron)
- Military systems:
  - Daimler-Chrysler

Power generation and distribution
- Courtesy of General Electric
- Courtesy of Kuka Robotics Corp.

Big Data is necessary for aML!

Does this all work here as well?
Medical Decision Making as a Search Task in \( H \)

Problem: Time \( (t) \)
Search in an arbitrarily high-dimensional space < 5 min.
06 iML
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,
  - Protein-Folding,
  - k-Anonymization,
  - Category Discovery, etc. etc....
Sometimes we (still) need a human-in-the-loop
Sometimes we need a doctor-in-the-loop
interactive Machine Learning (iML) := algorithms which interact with agents*) and can optimize their learning behaviour through this interaction

*) where the agents can be human


A group of experts-in-the-loop
A crowd of people-in-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** [1]
- **Example 2: k-Anonymization** [2]
- **Example 3: Protein Design** [3]


Example 3: Proteins are the building blocks of life ...

Protein prediction is an old – still unsolved - problem

The sequence of a protein can NOT (yet) be used to predict its 3D structure ...

Protein Design is a TSP problem


Hypothesis: most biological functions involve the interactions between many proteins, and the complexity of living systems arises as a result of such interactions.

In this context, the problem of inferring a global protein network for a given organism, - using all (genomic) data of the organism, is one of the grand challenges in computational biology.

Problem: Is Graph Isomorphism NP-complete?


- Important for health informatics: Discovering relationships between biological components
- Unsolved problem in computer science:
- Can the graph isomorphism problem be solved in polynomial time?
  - So far, no polynomial time algorithm is known.
  - It is also not known if it is NP-complete
  - We know that subgraph-isomorphism is NP-complete
Example: Holzinger Group Project: iML

http://hci-kdd.org/project/iML/
Reasons why ants find the shortest path (minimum linking model):

- 1) Earlier pheromones (the trail is completed earlier)
- 2) More pheromone (higher ant density)
- 3) Younger pheromone (less diffusion)

Soon, the ants will find the shortest path between their home and the food

No Free Lunch for the Ants

What is the probability for selecting a particular path?

\[ p_{ij} = \frac{[\tau_{ij}]^\alpha \cdot [\eta_{ij}]^\beta}{\sum_{l \in J_i} [\tau(t)]^\alpha \cdot [\eta]^\beta} \]

- \( p_{ij} \) is the probability of ants that they, at a particular node \( i \), select the route from node \( i \to j \) ("heuristic desirability")
- \( \alpha > 0 \) and \( \beta > 0 \) are the influence parameters (\( \alpha \) is the history coefficient, \( \beta \) the heuristic coefficient) usually \( \alpha \approx \beta \approx 2 < 5 \)
- \( \tau_{ij} \) is the pheromone value for the components, i.e. the amount of pheromone on edge \( (i,j) \)
- \( k \) is the set of usable components
- \( J_i \) is the set of nodes that ant \( k \) can reach from \( v_i \) (tabu list)
- \( \eta_{ij} = \frac{1}{d_{ij}} \) is the attractiveness computed by a heuristic, indicating the "a-priori desirability" of the move
Step 1: Pheromone update

The pheromone on each edge is updated as:

$$\tau_{ij} = (1 - \rho) \cdot \tau_{ij} + \Delta \tau_{ij}$$

With:

- $\rho$: the evaporation rate of the ‘old’ pheromone
- $\Delta \tau_{ij}$: the ‘new’ pheromone that is deposited by all ants on edge (i,j) calculated as:

$$\Delta \tau_{ij} = \sum_{k=0}^{m} \Delta \tau_{ij}^k$$
Step 2: Pheromone update

The pheromone that is deposited on edge \((i,j)\) by ant \(k\) is calculated as:

\[
\Delta \tau^k_{ij} = \begin{cases} 
\frac{Q}{L_k} & \text{if } (i, j) \in T_k \\
0 & \text{otherwise}
\end{cases}
\]

With:

- \(Q\) : a heuristic parameter
- \(T_k\) : the path traversed by ant \(k\)
- \(L_k\) : the length of \(T_k\) calculated as the sum of the lengths of all the edges of \(T_k\)
Input : ProblemSize, m, β, ρ, σ, q₀
Output: P_{best}

P_{best} ← CreateHeuristicSolution(ProblemSize);
P_{best\_cost} ← Cost(P_{best});

Pheromone_{init} ← \frac{1.0}{\text{ProblemSize} \times P_{best\_cost}};
Pheromone ← InitializePheromone(Pheromone_{init});

while ¬StopCondition() do
    for i = 1 to m do
        S_i ← ConstructSolution(Pheromone, ProblemSize, β, q₀);
        S_{i\_cost} ← Cost(S_i);
        if S_{i\_cost} ≤ P_{best\_cost} then
            P_{best\_cost} ← S_{i\_cost};
            P_{best} ← S_i;
        end
        LocalUpdateAndDecayPheromone(Pheromone, S_i, S_{i\_cost}, ρ);
    end
    GlobalUpdateAndDecayPheromone(Pheromone, P_{best}, P_{best\_cost}, ρ);
    while isUserInteraction() do
        GlobalAddAndRemovePheromone(Pheromone, P_{best}, P_{best\_cost}, ρ);
    end
end

return P_{best};

The attractiveness $\eta_{ij}$ of edge $(i, j)$ is computed by a heuristic, indicating the a-priori desirability of that particular move.

The pheromone trail level $\tau_{ij}$ of edge $(i, j)$ indicates how proficient it was in the past.

$\alpha = 0$ is a greedy approach and $\beta = 0$ represents the selection of tours that may not be optimal.

Consequently, we speak of a "trade-off" between speed and quality.
Evidence for the human-in-the-loop, however, strongly dependent on end-user expertise;

Further experiment will be camouflaged as a game, to provide the same start criteria for everybody.

You can have a look at the online Experiment:
http://hci-kdd.org/projects/iml-experiment
07 Conclusion
Successful Machine Learning needs a concerted effort international without boundaries ...
Application of aML in complex domains seems elusive in the near future, a good example are Gaussian processes, where aML (e.g., kernel machines) struggle on function extrapolation - which is trivial for human learners.

Making use of human cognitive abilities can be used to solve problems, when lacking big data sets, on complex data and/or rare events, where traditional learning algorithms suffer due to insufficient training samples.

Here human expertise can assist in solving problems which otherwise would remain NP-hard.

Successful application of ML requires a concerted effort of (1) data science, (2) algorithms, (3) graphs, (4) topology, (5) entropy, (6) data visualization, and (7) privacy, data protection, safety, and security—following the HCI-KDD approach in combining the best of two worlds.
08 Appendix


