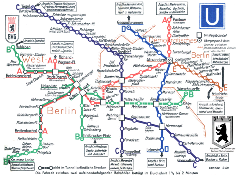
Optimization meets Machine Learning

Marco Lübbecke
Lehrstuhl für Operations Research
RWTH Aachen University

@mluebbecke

OR 2017 · Berlin · September 6, 2017

Optimization is Everywhere



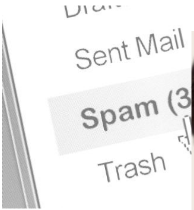
Machine Learning is Everywhere

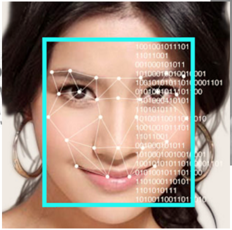
Machine Learning is Everywhere

Machine Learning is Everywhere

Machine Learning is Everywhere

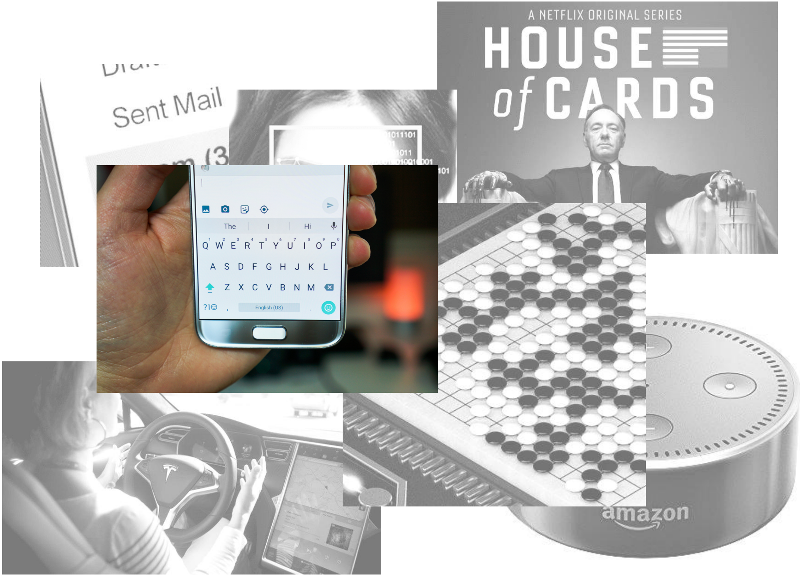
Machine Learning is Everywhere





Machine Learning is Everywhere

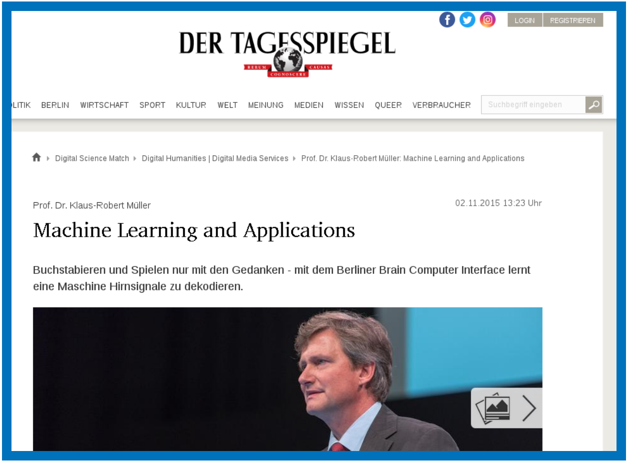
Machine Learning is Everywhere



Machine Learning is Everywhere

Machine Learning is Everywhere

Literally *Everyone* speaks about Machine Learning



DER TAGESSPIEGEL

LOGIN REGISTRIEREN

LITIK BERLIN WIRTSCHAFT SPORT KULTUR WELT MEINUNG MEDIEN WISSEN QUEER VERBRAUCHER

Suchbegriff eingeben

Digital Science Match › Digital Humanities | Digital Media Services › Prof. Dr. Klaus-Robert Müller: Machine Learning and Applications

Prof. Dr. Klaus-Robert Müller

02.11.2015 13:23 Uhr

Machine Learning and Applications

Buchstabieren und Spielen nur mit den Gedanken - mit dem Berliner Brain Computer Interface lernt eine Maschine Hirnsignale zu dekodieren.

Literally *Everyone* speaks about Machine Learning

DER TAGESSPIEGEL
LOGIN REGISTRIEREN
LITIK BERLIN
Digital
Prof. Dr. Kl
Mach
Buchstab
eine Mas

ABO SHOP AKADEMIE JOBS MEHR ▾ E-PAPER AUDIO APPS ARCHIV ANMELDEN

ZEIT ONLINE

Suche

Politik Gesellschaft Wirtschaft Kultur ▾ Wissen **Digital** Campus ▾ Karriere Entdecken Sport ZEITmagazin mehr ▾ #D17

Maschinelles Lernen

Künstlich, intelligent und frei

Firmen wie IBM, Facebook und Google investieren Millionen in künstliche Intelligenz. Teile ihrer Forschung stellen sie für alle frei zur Verfügung – aus gutem Grund.

Von **Eike Kühl**

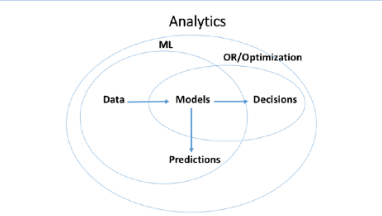
14. Dezember 2015, 19:28 Uhr / 19 Kommentare

Literally *Everyone* speaks about Machine Learning

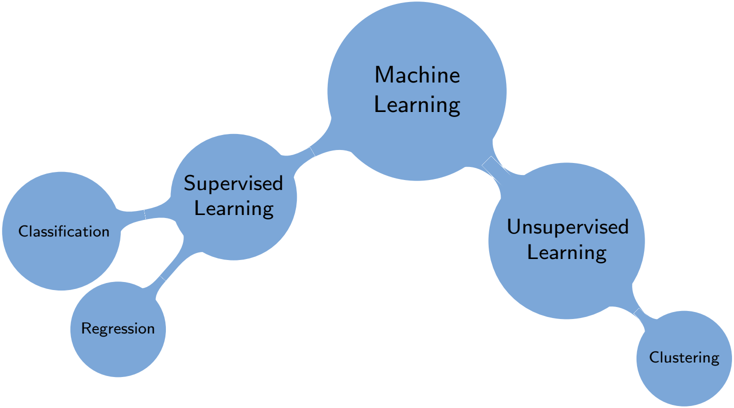
New INFORMS Journal on Optimization

> One of the largest opportunities of the field of optimization is to embrace data in a protagonist role and combine it with machine learning.
> — Dimitris Bertsimas, Editorial Statement

> My vision of the future for [. . .] optimization



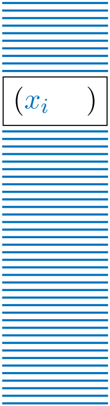
What *is* Machine Learning?



source: www.mathworks.com

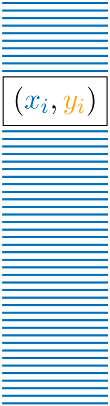
Supervised Learning: Classification

- data $\mathcal{X}$



Supervised Learning: Classification

- data $\mathcal{X}$ labels $\mathcal{Y}$



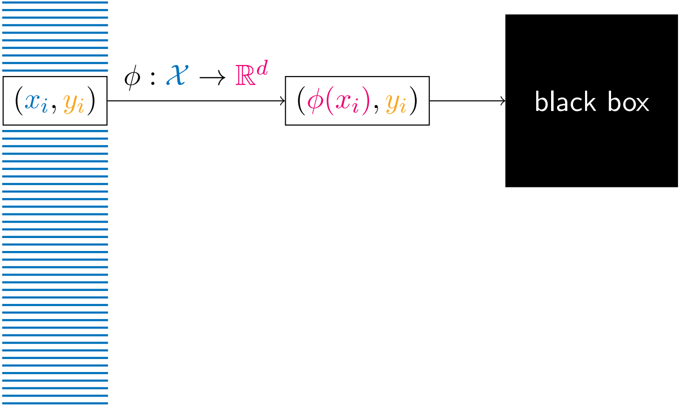
Supervised Learning: Classification

- data $\mathcal{X}$, d features, labels $\mathcal{Y}$

$$\boxed{(x_i, y_i)} \xrightarrow{\phi : \mathcal{X} \to \mathbb{R}^d} \boxed{(\phi(x_i), y_i)}$$

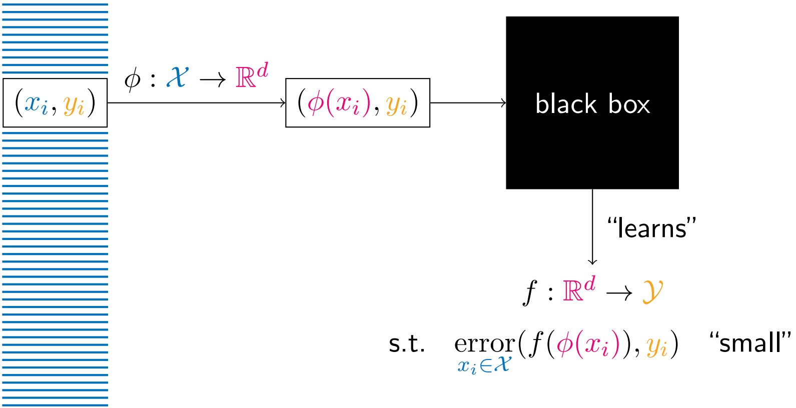
Supervised Learning: Classification

- data $\mathcal{X}$, d features, labels $\mathcal{Y}$



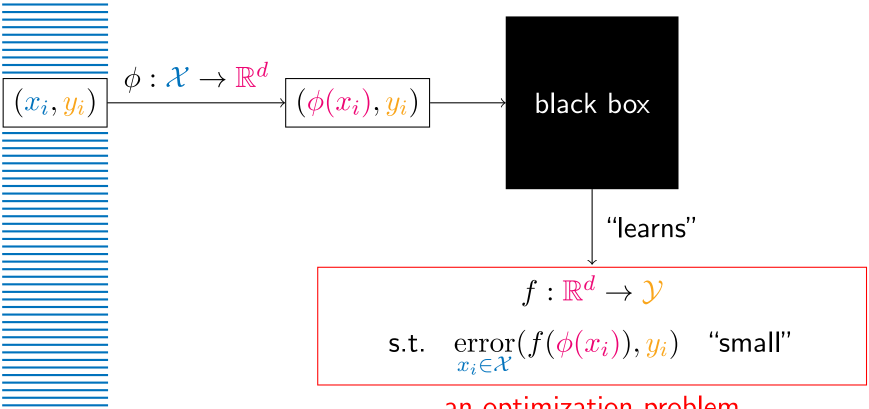
Supervised Learning: Classification

- data $\mathcal{X}$, d features, labels $\mathcal{Y}$



Supervised Learning: Classification

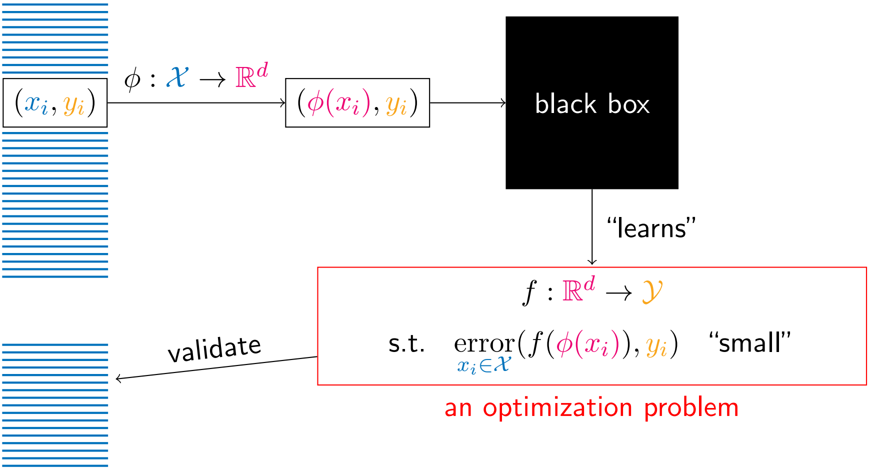
- data $\mathcal{X}$, d features, labels $\mathcal{Y}$



an optimization problem

Supervised Learning: Classification

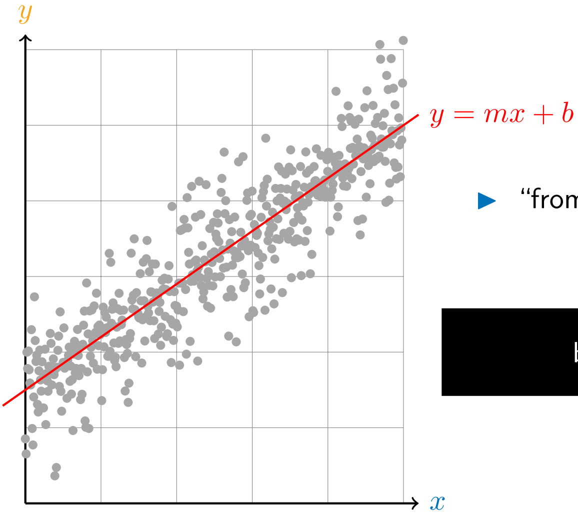
- data $\mathcal{X}$, d features, labels $\mathcal{Y}$



an optimization problem

Binary Classification: Dog or Muffin? Owl or Apple?

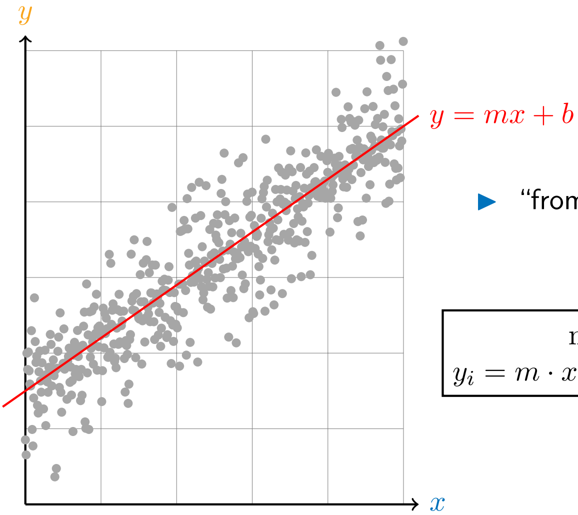
Supervised Learning: Regression



- “from $\{0,1\}$ to $[0,1]$”

black box

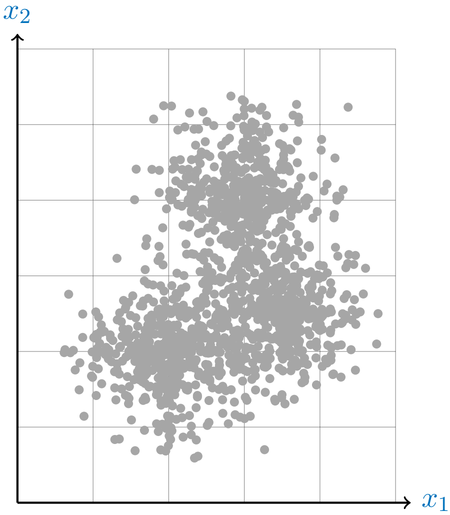
Supervised Learning: Regression

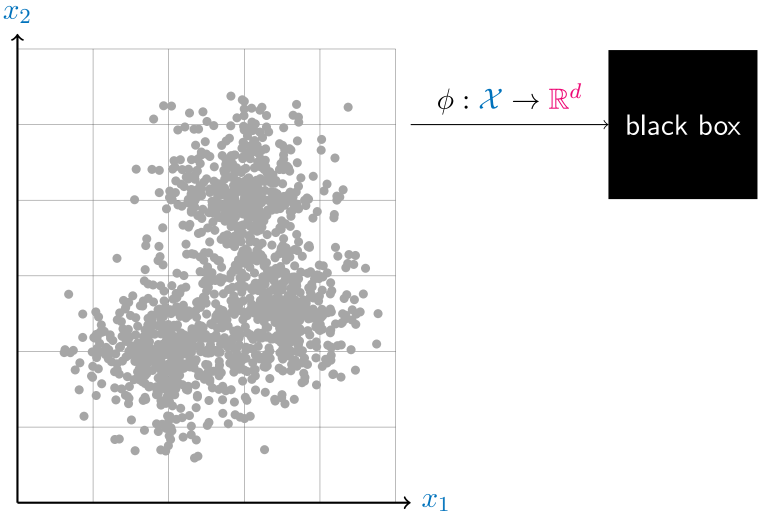


- "from $\{0, 1\}$ to $[0, 1]$"

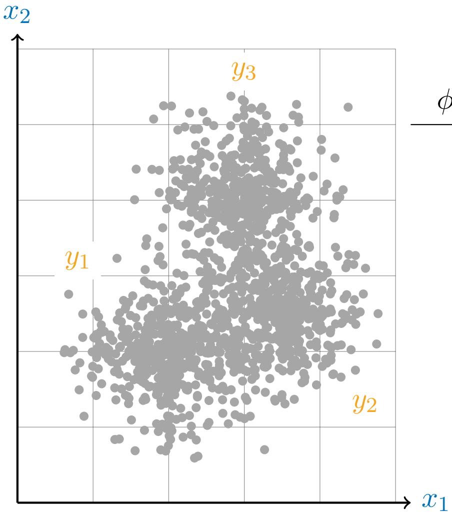
$$\boxed{\begin{array}{c} \min \sum_i \varepsilon_i^2 \\ y_i = m \cdot x_i + b + \varepsilon_i \quad x_i \in \mathcal{X} \end{array}}$$

Unsupervised Learning: Clustering




x_2
x_1
$\phi : \mathcal{X} \to \mathbb{R}^d$
black box

Unsupervised Learning: Clustering



$\phi : \mathcal{X} \to \mathbb{R}^d$

black box

"learns"

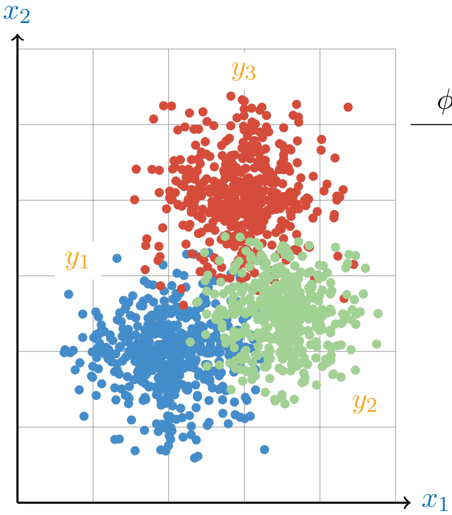
$f : \mathbb{R}^d \to \mathcal{Y}$

s.t. all

$x \in \mathcal{X} : f(\phi(x)) = y$

are "similar"

Unsupervised Learning: Clustering



$\phi : \mathcal{X} \to \mathbb{R}^d$

black box

"learns"

$f : \mathbb{R}^d \to \mathcal{Y}$

s.t. all

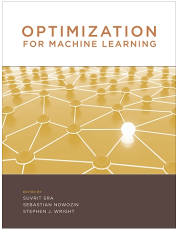
$x \in \mathcal{X} : f(\phi(x)) = y$

are "similar"

Optimization naturally appears in ML

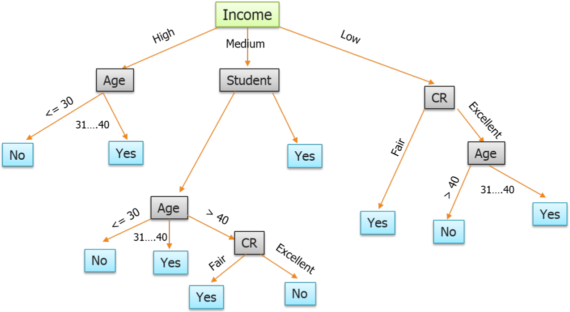
> Optimization lies at the heart of ML. Most ML problems reduce to optimization problems.
>
> — Bennett, Parrado-Hernández (2006)

- minimize e.g., prediction error
- continuous, convex optimization
- discrete, integer optimization



Example: A MIP in a black box : Classification Trees

- optimal classification trees Bertsimas & Dunn (2017)



source: www.edureka.co/blog/decision-trees

- use few nodes, shallow depth → formulated as MIP
- improves accuracy over classical CART method by 0.5–2%

Many Opportunities for Discrete Optimization in ML

- within black boxes to capture combinatorial explosion

→ see also Andrea's plenary on Friday

- feature selection
- outlier detection
- parameter tuning
- ...

How about the converse Direction?

- "emulating the expert" Bärmann, Pokutta & Schneider (2017)
- observe a decision maker $\max c_{\text{true}}^T x : x \in X(p)$
- learn their objective function given $(p_t, x_t^*)_{t=1,\dots,T}$
- *online learning*

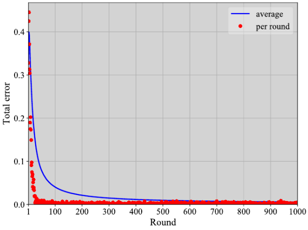


Figure 1. Linear Knapsack problem with $n = 100$ items over

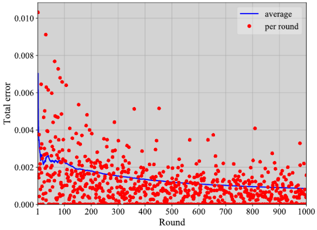


Figure 2. Resource-constrained shortest path problem on a grid

ML may help improving Optimization Algorithms

- e.g., branching in B&B
- *full strong branching* gives locally perfect information

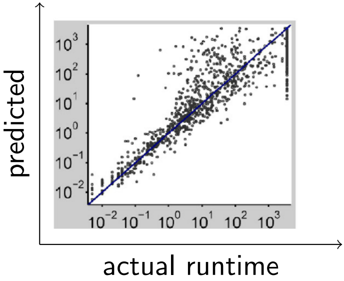
- predict the strong branching score of a variable
- features describe state of a variable
- supervised learning: regression

→ promising proof-of-concept
  Marcos Alvarez, Louveaux & Wehenkel (2017)

- survey on ML in branching/searching Lodi & Zarpellon (2017)

A Progress Bar for Branch-and-Bound?

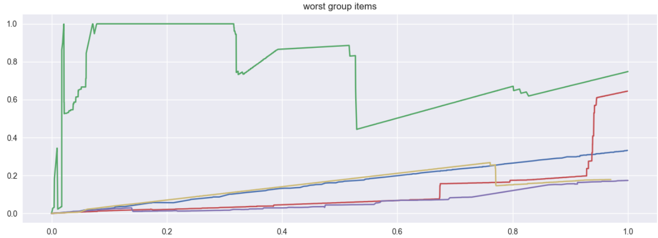
- predict the runtime of branch-and-bound algorithms
  Hutter, Xu, Hoos & Leyton-Brown (2014)
- CPLEX 12.1 on 1510 publicly available MIPs



A Progress Bar for Branch-and-Bound?

- very preliminary experiments with `gurobi` 7.5
- predict elapsed runtime percentage

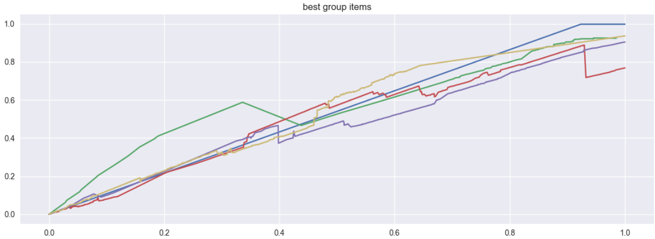
  Kruber, L, Obeloer genannt Bregenhorn (2017)



A Progress Bar for Branch-and-Bound?

- very preliminary experiments with `gurobi` 7.5
- predict elapsed runtime percentage

  Kruber, L, Obeloer genannt Bregenhorn (2017)



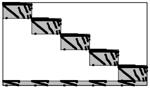
Learning when to solve a MIP by Branch-and-Price

- our MIP solver GCG detects many potential DW reformulations

⋮

MIP

+



CPU time

⋮

Kruber, L, Parmentier (2017)

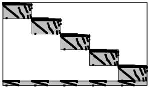
Learning when to solve a MIP by Branch-and-Price

- our MIP solver GCG detects many potential DW reformulations

⋮

MIP

+



CPU time

⋮

$\xrightarrow{\phi}$

100+ features

#conss, #vars, %constype, %vartype, #blocks, . . .

Kruber, L, Parmentier (2017)

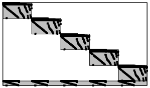
Learning when to solve a MIP by Branch-and-Price

- our MIP solver GCG detects many potential DW reformulations

⋮

MIP

+



CPU time

⋮

ϕ →

100+ features

#conss, #vars, %constype, %vartype, #blocks, . . .

→ k-NN

learns ↓

binary classifier f

"run SCIP or GCG?"

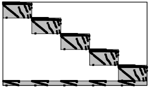
Kruber, L, Parmentier (2017)

- our MIP solver GCG detects many potential DW reformulations

⋮

MIP

+



CPU time

⋮

ϕ

100+ features
#conss, #vars, %constype, %vartype, #blocks, ...

k-NN

learns

binary classifier f

“run SCIP or GCG?”

faster

f		SCIP	GCG
	SCIP	69.5%	9.9%
	GCG	6.9%	13.7%

Kruber, L, Parmentier (2017)

What ML Answers can we (Optimizers) expect?

- we get statistical answers $\rightarrow$ not what we are used to see

- we have domain/expert knowledge: e.g., pseudo-costs
- ML may give a better predictor, but no *explanation*
- some info can be extracted from most influential features

$\rightarrow$ interpretability is a huge theoretical and practical topic

Decision Making: Machine Learning

> Machine Learning and Artificial Intelligence delivers the most value when you need to make lots of similar decisions quickly.
>
> — Ingo Mierswa, Rapidminer

- simple decisions: e.g., auto correct current word
- solution: often a single score $\rightarrow$ greedy
- keep/learn habits: extrapolate from the past (!)

Operations Research | RWTH AACHEN UNIVERSITY

Typical Example: Predictive Maintenance

source: blog.capterra.com/should-you-invest-in-a-predictive-maintenance-strategy/

Exploit all Options: Prescriptive Maintenance

source: www.siemens.com/press/pool/de/pressebilder/photonews/pn200826/300dpi/pn200826-12_300dpi.jpg

Decision Making: Optimization

> How often can the result of an optimization model be captured in a single variable?
>
> — Ed Rothberg, Gurobi

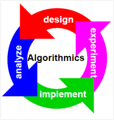
- solution: not only the objective value!
- complex decisions/plans: e.g., timetables, crew schedules, ...
- global scope: models all (reasonable) interdependencies

Perfect Partners

"Current Standard:" Predictive then Prescriptive Analytics

ML harnesses the bigness of data (the past and present);
Optimization captures the bigness of options (the future).

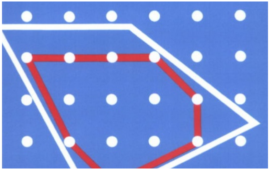
Learning (about) optimal Solutions

- in recurring complex decision situations
- learn how good (partial) solutions look like
- this may help finding good solutions faster
- learn spatio-temporal patterns to generate effective schedules
  Le, Liu, Lau (2016)
- 

Learning (about) Optimization Models

- ML can make sense of data
- optimization models are also “data”

→ (how) can ML help us make sense of optimization models?



- can ML learn good modeling?

Vision: Learning (about) Optimization Problems

- learn the *semantics* of a MIP model ("the problem")

$$
\begin{aligned}
\min \sum_{j=1}^{n} y_j & & \\
\text{s.t.} \sum_{j=1}^{n} x_{ij} &= 1 & i = 1,\dots,n \\
\sum_{i=1}^{n} a_i x_{ij} &\leq b & j = 1,\dots,n \\
x_{ij} &\leq y_j & i,j = 1,\dots,n \\
x_{ij},\, y_j &\in \{0,1\} & i,j = 1,\dots,n
\end{aligned}
$$

$\longrightarrow$

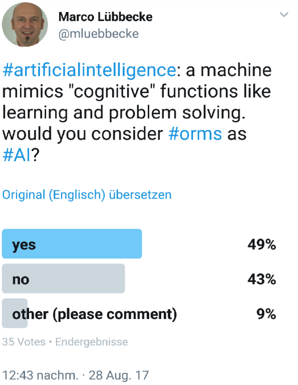
$\longrightarrow$

$$
\begin{aligned}
\min \sum_{p \in P} \nu_p & & \\
\text{s.t.} \sum_{p \in P : i \in p} \nu_p &= 1 & i = 1,\dots,n \\
\nu_p &\in \{0,1\} & p \in P
\end{aligned}
$$

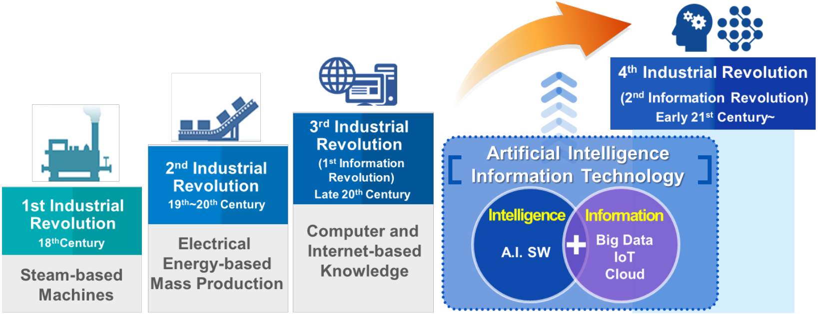
$\Rightarrow$ e.g., help the modeler find a better formulation

The AI Umbrella?

- OR vs. analytics discussion ✓
- OR vs. AI discussion ?



Why is this Relevant?



source: blogs.worldbank.org/category/tags/artificial-intelligence

- *if* the fourth industrial revolution is about AI, OR should be part of it

Optimization met Machine Learning

Marco Lübbecke
Lehrstuhl für Operations Research
RWTH Aachen University

@mluebbecke

OR 2017 · Berlin · September 6, 2017