Machine Learning and the AI thread

Michèle Sebag

TAO

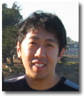
ECAI 2012, Turing session

cnrs
INRIA
UNIVERSITÉ PARIS-SUD 11
PASCAL
Pattern Analysis, Statistical Modelling and Computational Learning

Overview

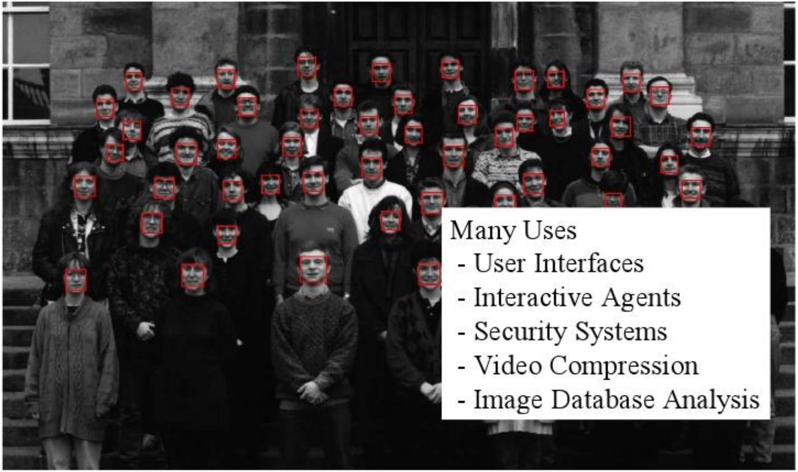
Examples

- Vision
- Control
- Netflix
- Spam
- Playing Go
- Google



http://ai.stanford.edu/∼ang/courses.html

Detecting faces

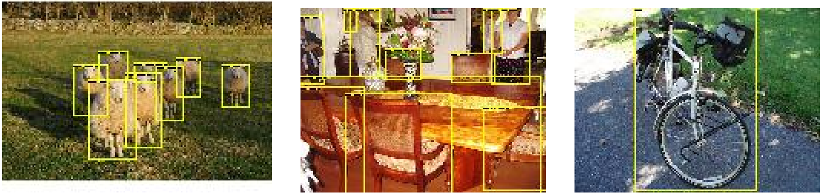


2

Viola and Jones, Robust object detection using a boosted cascade of simple features, CVPR 2001

The 2005-2012 Visual Object Challenges

A. Zisserman, C. Williams, M. Everingham, L. v.d. Gool



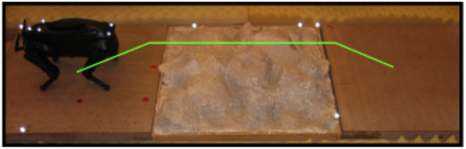
The 2005 Darpa Challenge

Thrun, Burgard and Fox 2005

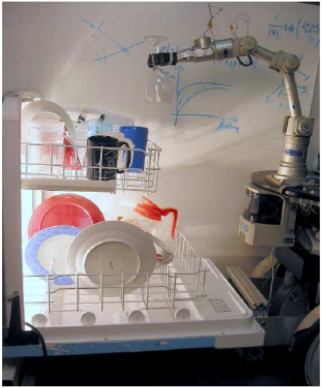
Autonomous vehicle Stanley – Terrains

Robots

Ng, Russell, Veloso, Abbeel, Peters, Schaal, ...



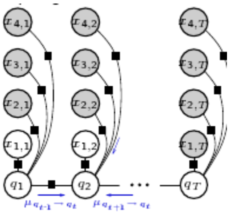
Reinforcement learning



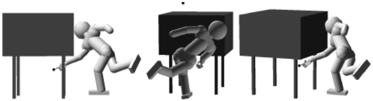
Classification

Robots, 2

Toussaint et al. 2010



(a) Factor graph modelling the variable interactions

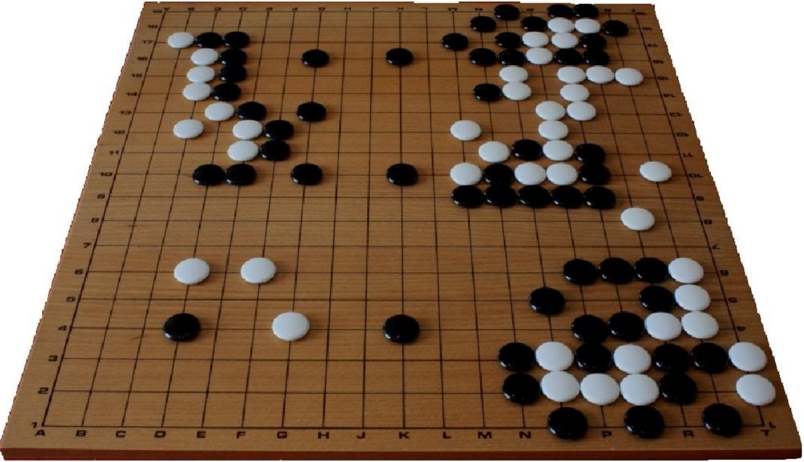


(b) Behaviour of the 39-DOF Humanoid:
Reaching goal under Balance and Collision constraints

Bayesian Inference for Motion Control and Planning

Go as AI Challenge

Gelly Wang 07; Teytaud et al. 2008-2011



Reinforcement Learning, Monte-Carlo Tree Search

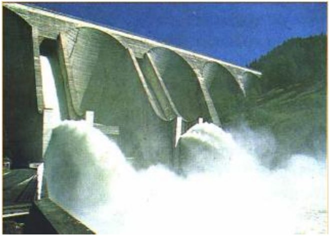
Energy policy

Claim

Many problems can be phrased as optimization in front of the uncertainty.

Adversarial setting	2 two-player game
uniform setting	a single player game

Management of energy stocks under uncertainty



Netflix Challenge 2007-2008

Collaborative Filtering

Spam – Phishing – Scam

Classification, Outlier detection

The power of big data

- Now-casting outbreak of flu
- Public relations >> Advertizing

Mc Luhan and Google

We shape our tools and afterwards our tools shape us

Marshall McLuhan, 1964

First time ever a tool is observed to modify human cognition that fast.

Sparrow et al., Science 2011

Overview

AI research agenda

J. McCarthy 56

We propose a study of artificial intelligence [..]. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.

Alan Turing

Machine Learning, 1950

by (...) mimicking education, we should hope to modify the machine until it could be relied on to produce definite reactions to certain commands.

Before AI...

Alan Turing

Machine Learning, 1950

by (...) mimicking education, we should hope to modify the machine until it could be relied on to produce definite reactions to certain commands.

How ?

One could *carry through the organization of an intelligent machine with only two interfering inputs, one for pleasure or reward, and the other for pain or punishment.*

The imitation game

The criterion:

Whether the machine could answer *questions in such a way that it will be extremely difficult to guess whether the answers are given by a man, or by the machine*

Critical issue

The extent we regard something as behaving in an intelligent manner is determined as much by our own state of mind and training, as by the properties of the object under consideration.

The imitation game, 2

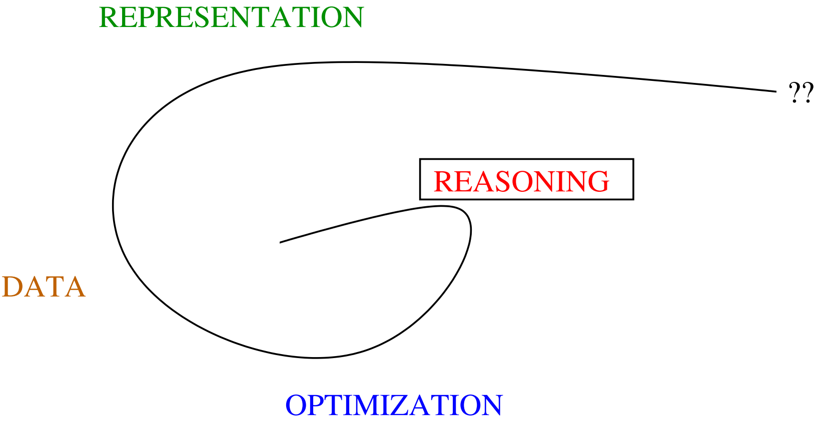
A regret-like criterion

- Comparison to reference performance (oracle)
- More difficult task $\not\Rightarrow$ higher regret

Oracle = human being

- Social intelligence matters
- Weaknesses are OK.

Overview


REPRESENTATION
??
REASONING
DATA
OPTIMIZATION

AI and ML, first era

General Problem Solver
. . . not social intelligence

Focus — Alan Bundy, wednesday

- Proof planning and induction
- Combining reasoners and theories

AM and Eurisko — Lenat 83, 01

- Generate new concepts
- Assess them

Lessons

Lenat 2001

the promise that *the more you know the more you can learn (..) sounds fine until you think about the inverse, namely, you do not start with very much in the system already. And there is not really that much that you can hope that it will learn completely cut off from the world*.

Interacting with the world is a must-have

The Robot Scientist

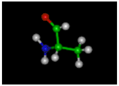
King et al, 04, 11

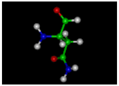
The robot scientist: completes the cycle
from hypothesis to experiment to reformulated hypothesis
without human intervention.

The Robot Scientist, 2

Why does it work ?

- A proper representation

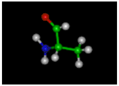


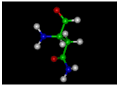


The Robot Scientist, 2

Why does it work ?

- A proper representation



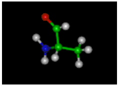


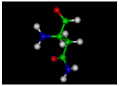
- Active Learning – Design of Experiment

The Robot Scientist, 2

Why does it work ?

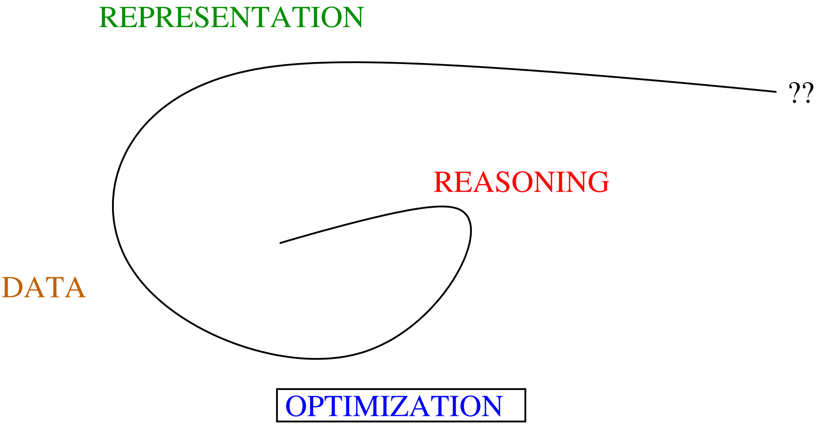
- A proper representation





- Active Learning – Design of Experiment
- Control of noise

Overview


REPRESENTATION
??
REASONING
DATA
OPTIMIZATION

ML second era: Optimization is everything

In neural nets

- Weights
- Structure

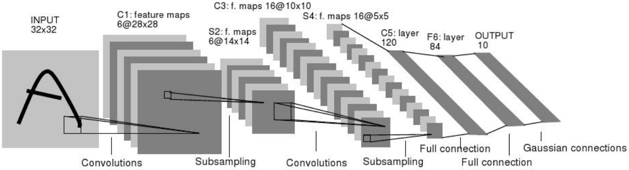
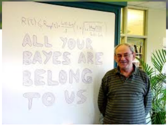


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

There has been several demonstrations that, with enough training data, learning algorithms are much better at building complex systems than humans: speech and hand-writing. Le Cun 86

Convex optimization is everything



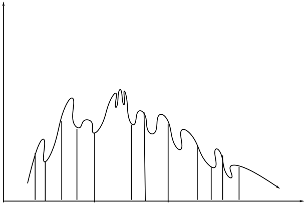
Goal: Minimize the loss

- On the training set: empirical error $\frac{1}{n}\sum_i \ell(h(x_i), y_i)$
- On the whole domain: generalization error $\int \ell(y, h(x))dP(x, y)$

Statistical machine learning

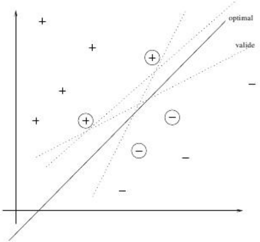
Vapnik 92, 95

Generalization error $<$ Empirical error
$+$ Regularity (h, n)



Support Vector Machines

Not all separating hyperplanes are equal



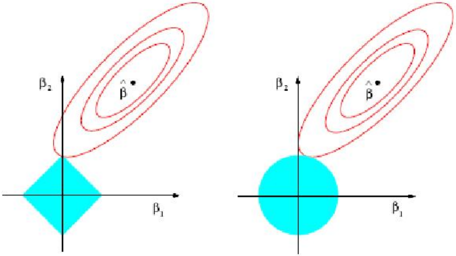
Divine surprise: a quadratic optimization problem

Boser et al. 92

$$\begin{cases} \text{Minimize} & \frac{1}{2}\,||w||^2 \\ \text{subject to} & \forall\ i,\ \ y_i(\langle w, x_i \rangle + b) \geq 1 \end{cases}$$

Tibshirani 96, Ng 04

Regularization term: parsimony and norm L_1



Use prior knowledge

Bach 04; Mairal et al. 10

- Given a structure on the features,
- ... use it within the regularization term.

Convex optimization, but ...

Achilles' heel

- Tuning hyper-parameters (regularization weight, kernel parameters): Cross-Validation

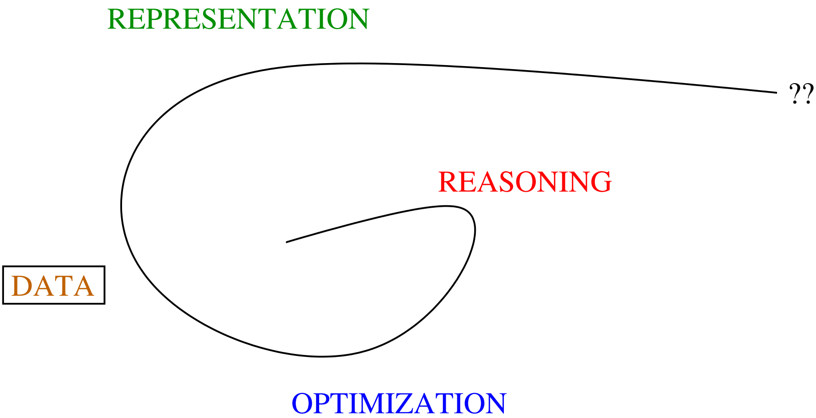
More generally

- Algorithm selection: Meta-learning — Bradzil 93

Much more generally

- Problem reduction — Langford 06

Overview


REPRESENTATION
??
REASONING
DATA
OPTIMIZATION

ML third era: all you need is more !

- More data
- More hypotheses
- (Does one still need reasoning ?)

All you need is more data

If algorithms are consistent — Daelemans 03

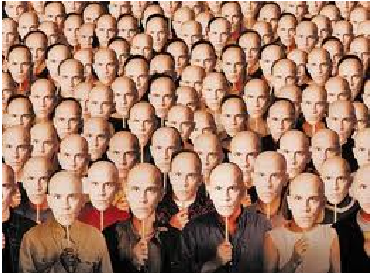
- When the data amount goes to infinity,
- ... all algorithms get same results

When data size matters

- Statistical machine translation
- The **textual entailment challenge** — Dagan et al. 05
  - Text: *Lyon is actually the gastronomic capital of France*
  - Hyp: *Lyon is the capital of France*
  - Does T entail H ?

Ensemble learning

- The strength of weak learnability — Schapire 90
- The wisdom of crowds



NO

YES

Random Forests

oldies but goodies

Example: KDD 2009 Challenge

1. Churn
2. Appetency
3. Up-selling

Is more data all we need ?

A thought experiment — Grefenstette, pers.

- The web: a world of information
- Question: what is the color of cherries ?

Is more data all we need ?

A thought experiment — Grefenstette, pers.

- The web: a world of information
- Question: what is the color of cherries ?
- After Google hits, 20% of cherries are black...

Is more data all we need ?

A thought experiment — Grefenstette, pers.

- The web: a world of information
- Question: what is the color of cherries ?
- After Google hits, 20% of cherries are black...

- Something else is needed...

Overview

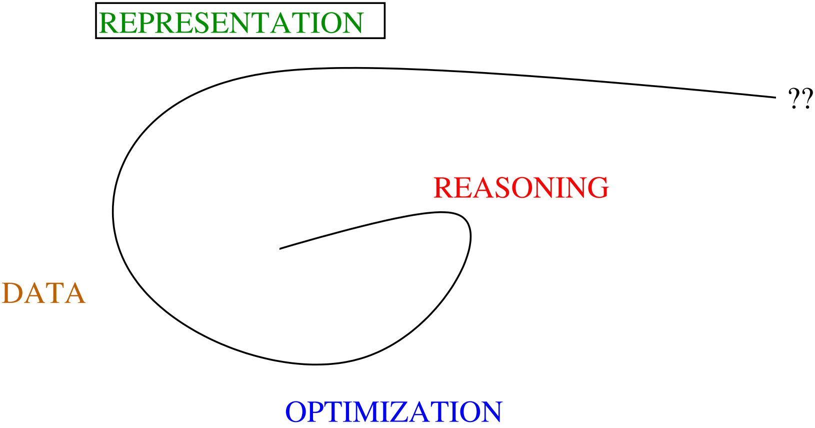
Some promises have been held

The initial vision

The spiral development of ML

- Reasoning
- Optimization
- Data
- Representation

Conclusion


REPRESENTATION
??
REASONING
DATA
OPTIMIZATION

Representation is everything

- Bayesian nets — Pearl 00
- Deep Networks — Hinton et al. 06, Bengio et al. 06
- Dictionary learning — Donoho et al. 05; Mairal et al. 10

Pearl 2000

- associational inference
  what if I see X ?
  evidential or statistical reasoning

Pearl 2000

- associational inference
  what if I see X ?
  evidential or statistical reasoning
- interventional inference
  what if I do X ?
  experimental or causal reasoning

Causality: Models, Reasoning and Inference

Pearl 2000

- associational inference
  what if I see X ?
  evidential or statistical reasoning
- interventional inference
  what if I do X ?
  experimental or causal reasoning
- retrospectional inference
  what if I had not done X ?
  counterfactual reasoning

Deep Networks

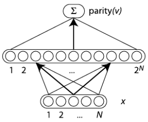
Hinton et al. 06, Bengio et al. 06

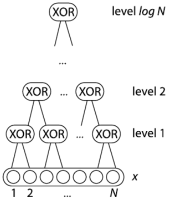
Grand goal

- Using ML to reach AI: (...) understanding of high-level abstractions
- Trade-off: computational, statistical, student-labor efficiency

Bottleneck

- Pattern matchers: partition the space
- Inefficient at representing highly varying functions





Greedy Learning of Multiple Levels of Abstractions

- Learning AI $\Rightarrow$ **learning abstractions**
- General principle: **Greedily learning simple things first, higher-level abstractions on top of lower-level ones.**
- **Implicit prior**: restrict to functions that
  1. can be represented as a composition of simpler ones such that
  2. the simpler ones can be learned first (i.e., are also good models of the data).
- Coherent with psychological literature (Piaget 1952).
  We learn baby math before arithmetic before algebra before differential equations . . .
- Also some evidence from neurobiology: (Guillery 2005) *"Is postnatal neocortical maturation hierarchical?"*.

Dictionary Learning

Principle

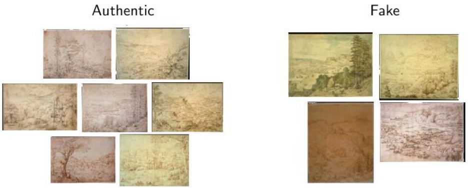
- A large dictionary, where you can express your thoughts in few words
- Robustness against noise

Dictionary Learning

Principle

- A large dictionary, where you can express your thoughts in few words
- Robustness against noise

Hugues et al. 09; Mairal et al. 10



Dictionary Learning

Principle

- A large dictionary, where you can express your thoughts in few words
- Robustness against noise

Hugues et al. 09; Mairal et al. 10

Dictionary Learning

Principle

- A large dictionary, where you can express your thoughts in few words
- Robustness against noise

Hugues et al. 09; Mairal et al. 10

Overview

Conclusion

- Reasoning, Optimization, Data, Representation needed
- (Lifelong learning likely necessary)
- Prior knowledge needed

... one could carry through the organization of an intelligent machine with only two interfering inputs, one for pleasure or reward, and the other for pain or punishment...

What is needed:

- Prior knowledge or reward ?

Inspiration from a neighbor field

Human competitive (Humies) award

GECCO 2012

- Yavalath: an automatically designed game
- more popular than Backgammon and Chinese Checkers

What was the optimization objective ?

C. Browne

uncertainty; killer moves; permanence; completion; duration (negative)

Inspiration from a neighbor field

Human competitive (Humies) award

GECCO 2012

- Yavalath: an automatically designed game
- more popular than Backgammon and Chinese Checkers

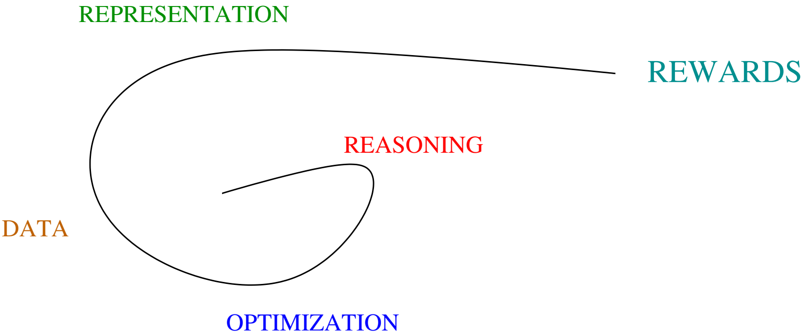
What was the optimization objective ?

C. Browne

uncertainty; killer moves; permanence; completion; duration (negative)

Then what should an AI system learn ?

Learn the objective


REPRESENTATION
REWARDS
REASONING
DATA
OPTIMIZATION