Learning Machines for Clinical Applications

Magnus Boman
Backdrop
Backdrop

- Two projects with KI (PI: Dr Viktor Kaldo)  
  - Vetenskapsrådet  
  - Erling-Perssons Stiftelse (split with SICS)

- Student opportunity  
  - New PhD student, Negar Safinianaini  
  - M Sc students  
  - Learning Machines course (IK3616/ID2225)

- Research collaboration opportunity  
  - One of the first Learning Machines projects ever
Intelligent Machinery, A Heretical Theory*

A. M. Turing

‘You cannot make a machine to think for you.’ This is a commonplace that is usually accepted without question. It will be the purpose of this paper to question it.
If the machine were able in some way to ‘learn by experience’ it would be much more impressive. If this were the case there seems to be no real reason why one should not start from a comparatively simple machine, and, by subjecting it to a suitable range of ‘experience’ transform it into one which was much more elaborate, and was able to deal with a far greater range of contingencies. This process could probably be hastened by a suitable selection of the experiences to which it was subjected. This might be called ‘education’.
Abstract

One long-term goal of machine learning research is to produce methods that are applicable to highly complex tasks, such as perception (vision, audition), reasoning, intelligent control, and other artificially intelligent behaviors. We argue that in order to progress toward this goal, the Machine Learning community must endeavor to discover algorithms that can learn highly complex functions, with minimal need for prior knowledge, and with minimal human intervention. We present mathematical and empirical evidence suggesting that many popular approaches to non-parametric learning, particularly kernel methods, are fundamentally limited in their ability to learn complex high-dimensional functions. Our analysis focuses on two problems. First, kernel machines are shallow architectures, in which one large layer of simple template matchers is followed by a single layer of trainable coefficients. We argue that shallow architectures can be very inefficient in terms of required number of computational elements and examples. Second, we analyze a limitation of kernel machines with a local kernel, linked to the curse of dimensionality, that applies to supervised, unsupervised (manifold learning) and semi-supervised kernel machines. Using empirical results on invariant image recognition tasks, kernel methods are compared with deep architectures, in which lower-level features or concepts are progressively combined into more abstract and higher-level representations. We argue that deep architectures have the potential to generalize in non-local ways, i.e., beyond immediate neighbors, and that this is crucial in order to make progress on the kind of complex tasks required for artificial intelligence.
Scaling Learning Algorithms towards AI

Yoshua Bengio (1) and Yann LeCun (2)

(1) Yoshua.Bengio@umontreal.ca
Département d’Informatique et Recherche Opérationnelle
Université de Montréal,
(2) yann@cs.nyu.edu
The Courant Institute of Mathematical Sciences,
New York University, New York, NY

Abstract

One long-term goal of machine learning research is to produce methods that are applicable to highly complex tasks, such as perception (vision, audition), reasoning, intelligent control, and other artificially intelligent behaviors. We argue that in order to progress toward this goal, the Machine Learning community must endeavor to discover algorithms that can learn highly complex functions, with minimal need for prior knowledge, and with minimal human intervention. We present mathematical and empirical evidence suggesting that many popular approaches to non-parametric learning, particularly kernel methods, are fundamentally limited in their ability to learn complex high-dimensional functions. Our analysis focuses on two problems. First, kernel machines are shallow architectures, in which one large layer of simple template matchers is followed by a single layer of trainable coefficients. We argue that shallow architectures can be very inefficient in terms of required number of computational elements and examples. Second, we analyze a limitation of kernel machines with a local kernel, linked to the curse of dimensionality, that applies to supervised, unsupervised (manifold learning) and semi-supervised kernel machines. Using empirical results on invariant image recognition tasks, kernel methods are compared with deep architectures, in which lower-level features or concepts are progressively combined into more abstract and higher-level representations. We argue that deep architectures have the potential to generalize in non-local ways, i.e., beyond immediate neighbors, and that this is crucial in order to make progress on the kind of complex tasks required for artificial intelligence.
AI for Scientific Use (2000)

Use your KTH access to Springer-Link and download the PDF for free!
THE NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN

National Science and Technology Council

Networking and Information Technology Research and Development Subcommittee

October 2016
Deep learning

Reinforcement learning

Approximate learning

III. Learning Systems

Summary—In order to solve a new problem, one should first try using methods similar to those that have worked on similar problems. To implement this “basic learning heuristic” one must generalize on past experience, and one way to do this is to use success-reinforced decision models. These learning systems are shown to be averaging devices. Using devices which learn also which events are associated with reinforcement, i.e., reward, we can build more autonomous “secondary reinforcement” systems. In applying such methods to complex problems, one encounters a serious difficulty—in distributing credit for success of a complex strategy among the many decisions that were involved. This problem can be managed by arranging for local reinforcement of partial goals within a hierarchy, and by grading the training sequence of problems to parallel a process of maturation of the machine’s resources.
A learning machine is an autonomous self-regulating open reasoning machine system that actively learns in an unsupervised and decentralised manner.
Learning Machine system overview

Perception
- Feature extraction, anomaly detection / attention, clustering, prediction, ...
- Deep Neural Network, ...
- Latent Dirichlet Allocation, ...
- Parametric / non-parametric Bayesian, anomaly detection, ...

Reasoning
- Higher-order concept extraction, inference, explanation, hypothesis generation, ...

Interaction
- Presentation, feedback, knowledge exchange, exploration, communication, visualisation, ...
- Natural language
- LM to LM

Data in motion
- Timely interaction
- Continuous adaptation
- Online learning
- Perception
- Reasoning
- Interaction

Resources
- Uncertainty
- Time
- Compute and resource management
- Libraries
- Machine learning
- Graph algorithms
- Data processing engines
- Flink
- Spark
- Yarn
- Mesos
- Storage and streams
- HDFS
- Kafka
- PaaS

Gillblad, SICS
Learning Machines as Artificial Therapists

Perception → Reasoning → Interaction

- Latent Dirichlet Allocation, ...
- Parametric / non-parametric Bayesian, anomaly detection, ...
- Deep Neural Network, ...
- Natural language...
- Video / images
- Text
- Sensor data...
- Feature extraction, anomaly detection / attention, clustering, prediction, ...
- Higher-order concept extraction, inference, explanation, hypothesis generation, ...
- Presentation, feedback, knowledge exchange, exploration, communication, visualisation, ...

Data in motion:
- Online learning
- Continuous adaptation
- Timely interaction

Learning Machine system overview:
- PaaS
- Flink
- Spark
- Yarn
- Mesos
- HDFS
- Kafka

Machine learning:
- Graph algorithms, ...
- Storage and streams
- Compute and resource management
- Data processing engines
- Libraries
- Action / response
- Perception
- Reasoning
- Interaction

Resources
- Uncertainty
- Time
Learning Machines as Artificial Therapists
Health Analytics Data Flows

Boman, Kruse
*Global Health Action*, May 2017
after Boman, Gillblad
*IEEE Big Data* 2014
Interactive Learning in LMs

- Multi-LM systems (M2M)
- Human-LM interaction
- Multi-LM systems with Human-LM interaction
Humans in Interactive LM Learning

- Multi-LM systems (M2M)
  Humans may be modelled (possibly as noise), and LM goals may be set by and for humans

- Human-LM interaction
  Humans must be modelled, LM adapting to human goals, possibly dynamically

- Multi-LM systems with Human-LM interaction
  Humans must be modelled, LM adapting to human and other LM goals, possibly dynamically
Prediction Task

- Predict, as early as possible, and with as high accuracy as possible, which patients in the programme are in need of more support.

- The programme is cognitive-based therapy (CBT) for Internet-based psychiatry (IPSY)

- Possibly addressable as a churn problem
Churn Analysis

- Developed for Business Intelligence (i.e., turning enterprise raw data into valuable information) and corporate performance management

- The largest vendor of predictive analytics software (e.g., SPSS Modeler) sells churn analysis services, including ANNs, decision trees, regression modelling, genetic algorithms, text mining, clustering,...
Data Mining Basic Techniques for BI

• Classification
• Clustering
• Regression
• Event sequence analysis
• Forecasting

The Industry Lingo
Example: K-means Clustering of Churned Telco Customers

Cross-Disciplinary Work is Ongoing...


IPSY Data Available

- 5218 patients
- ~4500 variables
- Ethical permit granted
- Pre-processing state of treatment platform data (R scripts to clean data)
Examples of (Reductionist) Goals

• A conceptual model of the information flow in the IPSY programme

• Health data analysis flows, resulting in multi-dimensional descriptions of the style of the patient, the therapist, and the interaction between them which could be further used in other analyses, such as NLP

• Anonymity and synthetic populations for data sharing
Dichotomised Target Event for Treatment Outcome

1. Responders (decreasing symptoms significantly)

2. Remitters (moving from the clinical range to the normal range of symptoms)

3. Clinical significant change (being both Responder and Remitter)

4. Non-responders (no significant change)

5. Deteriorated (a significant worsening)

Cut-offs for each primary outcome measure will be used
Treatment Adherence

• Completed treatment modules
• Automatically logged platform activity
• Self-reported homework
• Registrations of exercises
Feature-Based Output Data from an ANN Pipeline for Emotion Recognition

Boman, Högman, Karali
Unpublished
Tack!
mab@kth.se