Automated Learning from Graph-Structured Data

Quanming Yao[1,2], Huan Zhao[2], Yongqi Zhang[2]

[1]Department of Electronic Engineering, Tsinghua University

[2]4Paradigm Inc.

https://quanmingyao.github.io/AutoML.github.io/acml21-tutorial.html

Tutorial Outline

1. What is Automated Machine Learning (AutoML)
   - Background on technical tools from machine learning
2. Automated Graph Neural Networks (GNN)
   - Explore neural architecture search for GNN
3. Automated Knowledge Graph (KG) Embedding
   - Explore AutoML for KG Embedding

Schedule at a Glance

Time	Event
00-45 minutes	**Part 1**: An introduction to Automated Machine Learning (AutoML)
	Speaker: Quanming Yao
45-90 minutes	**Part 2**: Automated Graph Neural Networks (GNN)
	Speaker: Huan Zhao
90-135 minutes	**Part 3**: Automated Knowledge Graph (KG) Embedding
	Speaker: Yongqi Zhang
135-150 minutes	**Part 4**: Discussion

Automated Recommender System (RecSys) Tutorial

Part 1: An Introduction to Automated Machine Learning (AutoML)

Quanming Yao[1,2]

[1]*Assistant professor - Department of Electronic Engineering, Tsinghua University*

[2]*Founding Leader (ML research team), 4Paradigm Inc.*

qyaoaa@tsinghua.edu.cn

Outline

1. What is Machine Learning?
2. What is Automated Machine Learning (AutoML)?
3. Summary & Next Works

What is Machine Learning (ML)?

Applications

Search Engine
Recommender Systems
Loss Assessment

Image Classification

Predict the class of the object

Security Monitoring
Bio-payment
Flow Statistics

Face Recognition

Who is the person

COVID Simulation
Petroleum Exploration
Drug Discovery

Drug Design

Learn to make decisions

Better Performance

Higher Efficiency

Definition

Structure　Samples

$$\min_x F(x; D)$$

Parameters

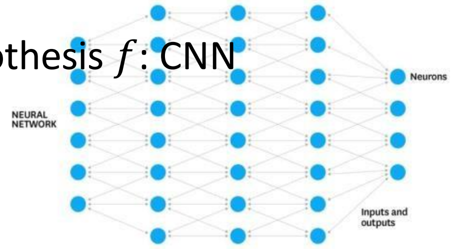
(iterative) optimization → Prediction Accuracy ↑

[1]. Machine Learning, Tom Mitchell, McGraw Hill, 1997.
[2]. 周志华著. 机器学习, 北京: 清华大学出版社, 2016年

ML = Data + Knowledge

Image Classification

Optimization ⇩



Hypothesis f: CNN

Generalization ⇩

Accuracy

Design a **hypothesis (function) f** to perform the learning task

Better Performance (more important) → Generalization → What kind of f should we use?

Higher Efficiency → Optimization → How can we find such f?

Target | Concept | Issue

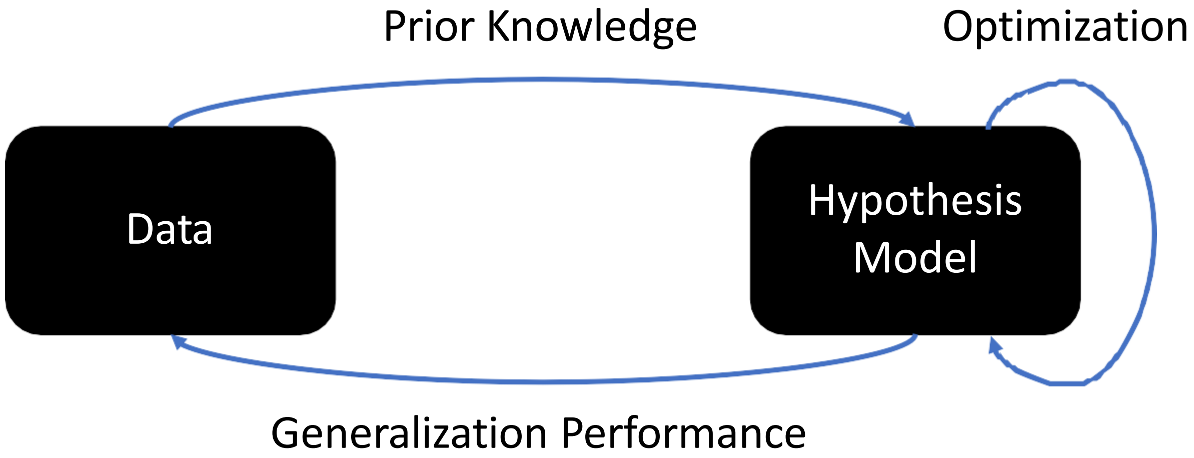
Not everything can be learnt

PAC-Learning (Definition 2.3 in [1]): What kind of problems can be solved in polynomial time

No Free Lunch Theorem (Appendix B [2]): No single algorithm can be good on all problems

[1]. M. Mohri, A. Rostamizadeh, A. Talwalkar. Foundations of machine learning. 2018
[2]. O. Bousquet, et.al. Introduction to Statistical Learning Theory. 2016

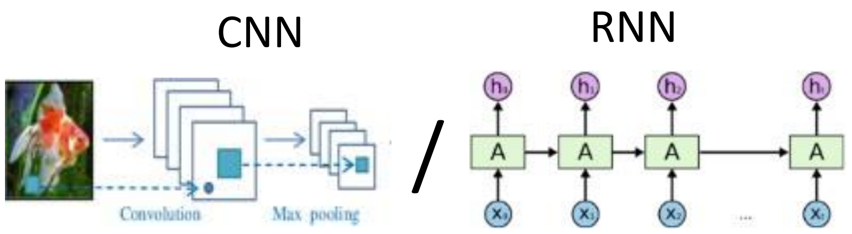
How to use ML Well?



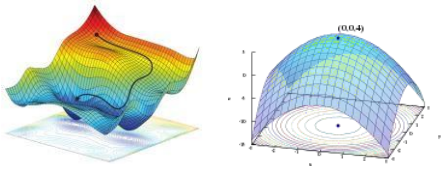
The Advancement of Learning

- An iteration between theory and practice

- A feedback loop



Generalization: What kind of f should we use?



SGD v.s. Adagrad[1]

Optimization: How can we find such f?

Prior knowledge

"All models are wrong, but some are useful"[2]

Better understanding of prior knowledge → Better hypothesis → Better generalization performance

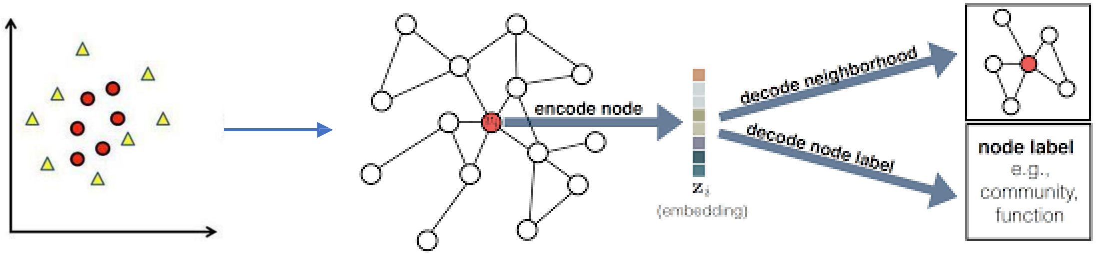
[1]. Image Source: A. Amini et al. "[Spatial Uncertainty Sampling for End-to-End Control]". NeurIPS Bayesian Deep Learning 2018
[2]. G. Box, Science and statistics, JASA 1976

Continual Trends in Machine Learning

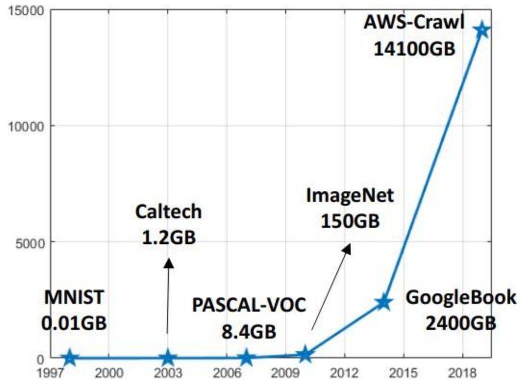
Given Task

$\min_{\boldsymbol{x}} F(x\,;D)$

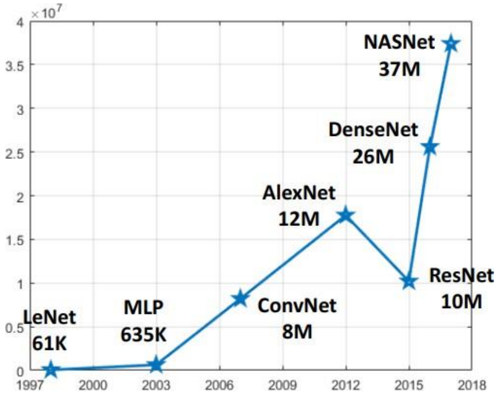
F: More Complicated Structure



D: More Samples



x: More Parameters



Moving Direction

Stronger models → better performance

Road Map in Recent History

- Core Issue in Machine Learning: Improving learning performance (with higher efficiency)

Rule-based

association rules mining,
learning classifier systems

1990s

Statistics-based

support vector machine,
sparse dictionary learning

2000s

Deep Learning-based

convolutional neural networks,
transformer

2010s

Better performance →

- Larger hypothesis (more complex models) are being used
- Optimization is getting complex (even mixed up with generalization)
- The prior knowledge is imposed on more abstract level

What is ML – Summary

- Machine learning = Data (optimization) + Knowledge (generalization)
  - Core Issue: Improving learning performance (with high efficiency)
- The advance and usage of ML is an iterative process
  - Better understanding of prior knowledge → Better generalization performance
- Continual trends in machine learning
  - The prior knowledge is imposed on more abstract level

Outline

1. What is Machine Learning?
2. What is Automated Machine Learning (AutoML)?
   - Explanation from a Simple Example
   - Recent Industrial and Research Examples
   - A Generalization Viewpoint for AutoML
   - How to use AutoML
3. Summary & Next Works

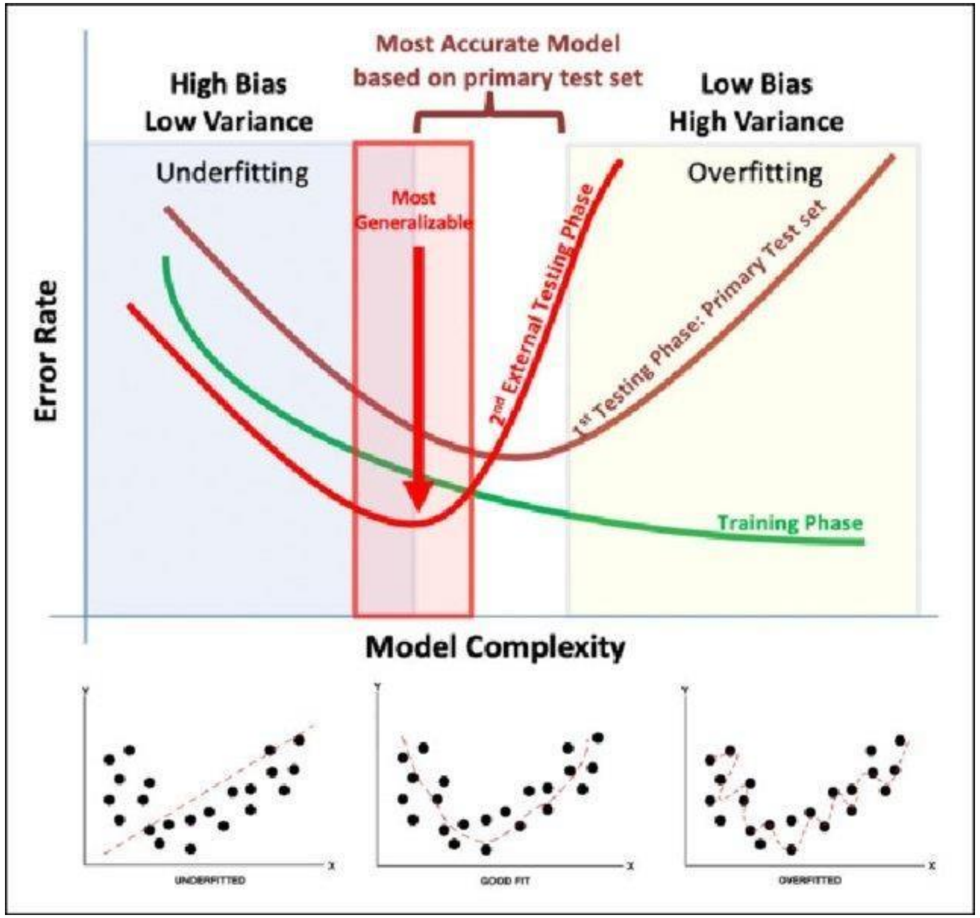
Simple Example – Tune hyper-parameter

Bi-level optimization

Hyper-parameter

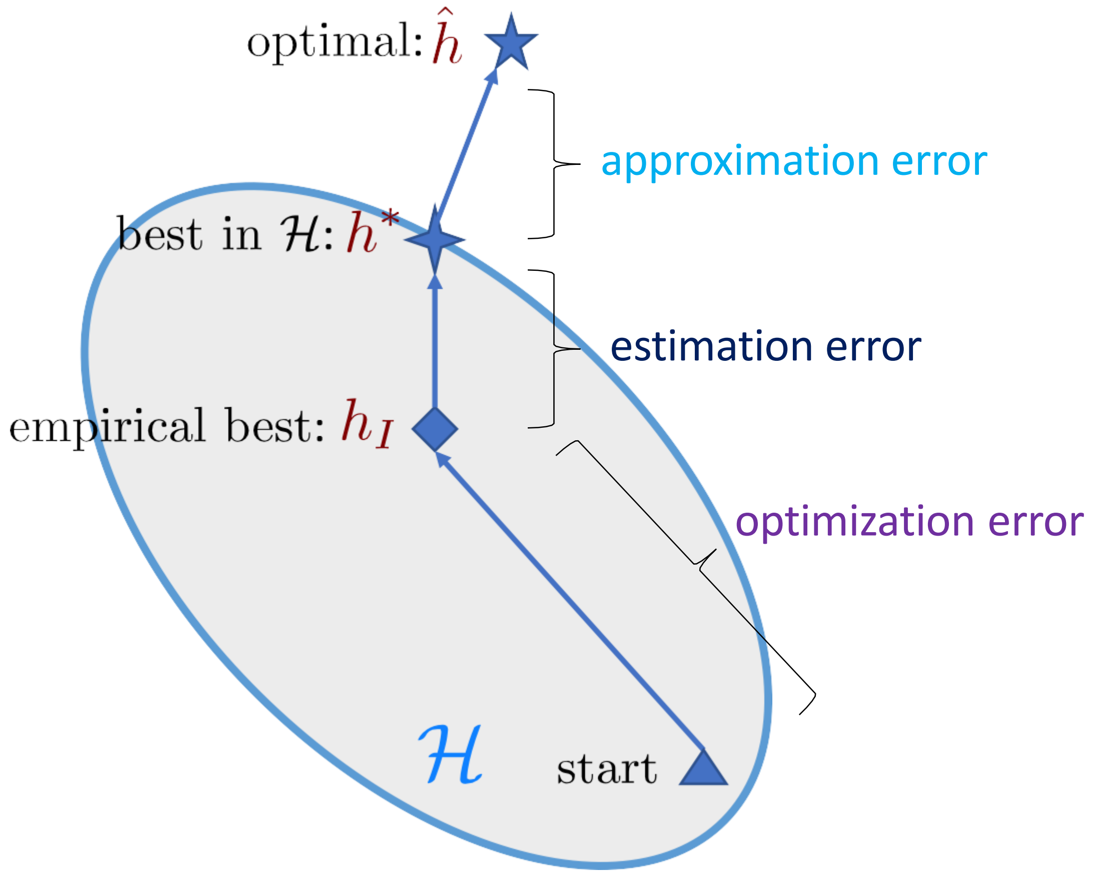
$$\underbrace{\max_{\lambda} \sum_{j} h(x_j; w^*)}_{\text{Validation Performance}} \quad \text{s.t.} \quad \underbrace{\min_{w} \sum_{i} f(x_i; w) + \lambda \|w\|_1}_{\text{Training objective}}$$

- Large λ leads to sparse w^*
- Grid search: enumerating $\lambda \in \{1,2,4,8, \ldots\}$



[1]. Image source: Artificial Intelligence and Machine Learning in Pathology: The Present Landscape of Supervised Methods.

Mach. Learn – Error decomposition



Total error in machine learning

- Approximation error
  - Which classifier to be used
  - What are their hyper-parameters
  - Distribution changes
- Estimation error
  - Finite samples
  - Regularization hyper-parameter
- Optimization error
  - Which algorithm to be used
  - How to tune its step-size

Reduce

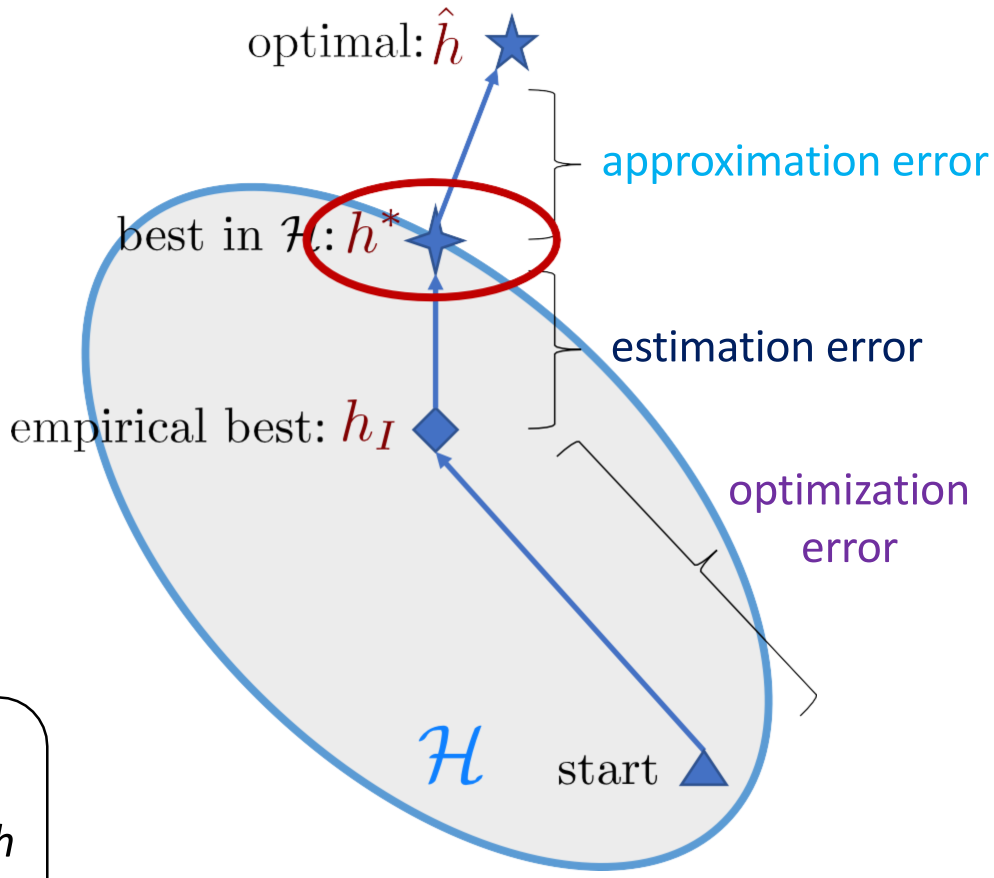
$$\min_w \sum_i f(x_i; w) + \lambda \|w\|_1$$

Image is from Y. Wang, et. al. Generalizing from a Few Examples: A Survey on Few-Shot Learning. CSUR 2020

Look Inside Error Decomposition

Automatically find h^* by bi-level optimization

$$\underbrace{\max_{\lambda} \sum_{j} h(x_j; w^*)}_{\text{Validation Performance}} \quad \text{s.t.} \quad \underbrace{\min_{w} \sum_{i} f(x_i; w) + \lambda \|w\|_1}_{\text{Training objective}}$$



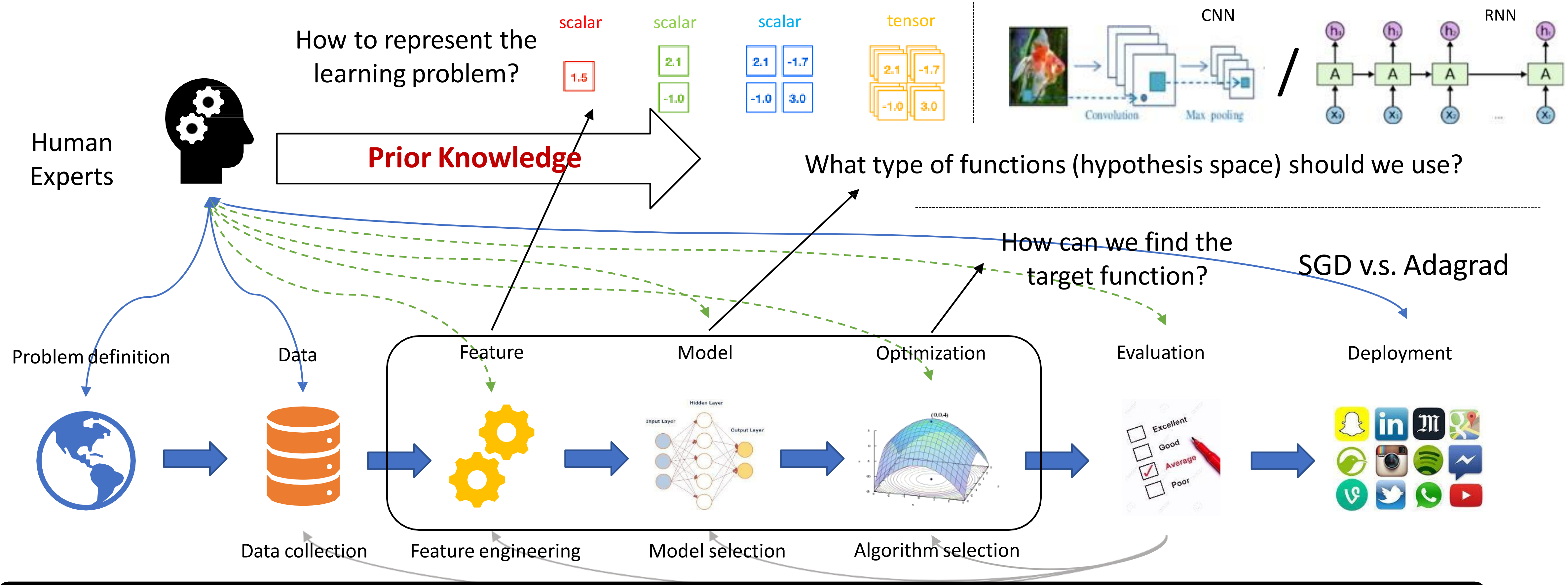
How to further improve the performance in an automatic manner (i.e., reduce the approximation error)?

- Feature can be weak → *Automatic feature engineering*
- Linear predictor can be too restrictive → *Neural architecture search*
- Grid search can be slow → *Search in a supernet*

AutoML

Figure is from Q. Yao et.al. Taking Human out of Learning Applications: A Survey on Automated Machine Learning. arXiv 2018

What is AutoML – Practical Viewpoint

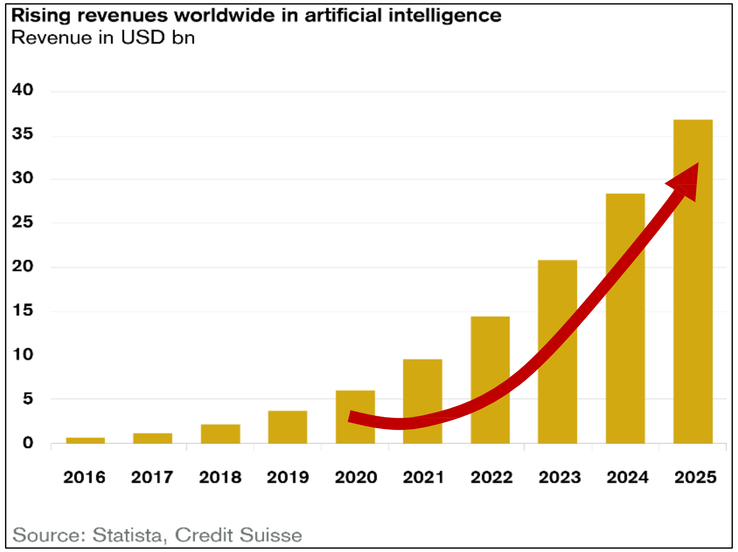


Parameterize (low-level) prior knowledge in the usage and design of machine learning

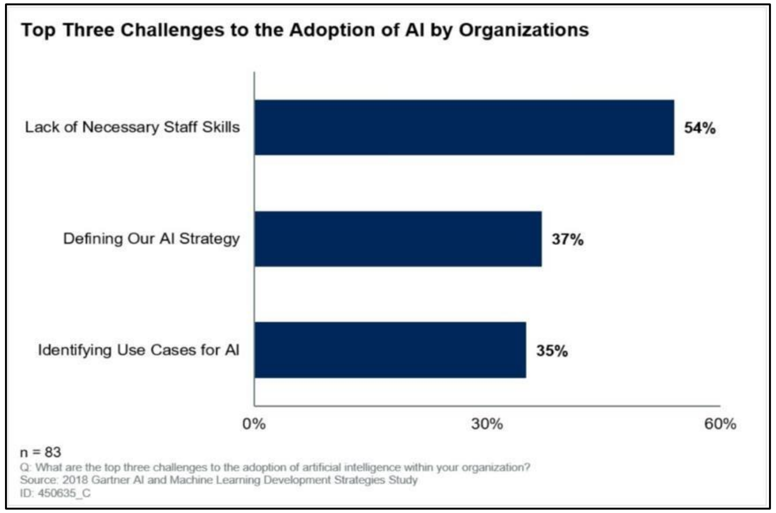
As a consequence

- Human participations can be naturally replaced by computation power
- total error of machine learning can be reduced (generalization can be improved)

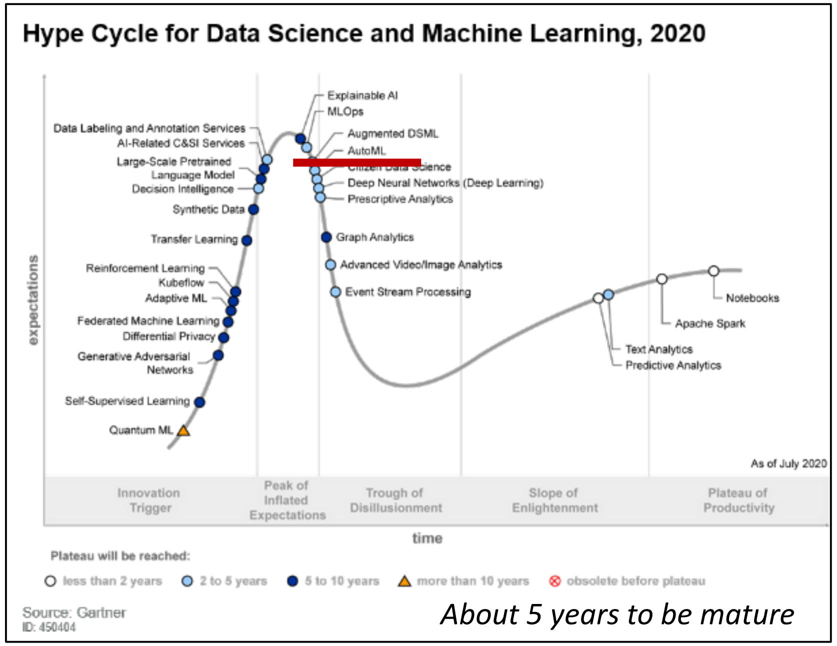
Why We need AutoML?



Investment in AI industry



Practical needs



Technical trends

- **Industry** – reduce the expense, increase usage coverage – huge market value [1]
- **Academy** – understanding data science on a higher level – great intelligence value [2,3]

[1]. Gartner: https://www.forbes.com/sites/janakirammsv/2020/03/02/key-takeaways-from-the-gartner-magic-quadrant-for-ai-developer-services/#a95b99ee3e5e
[2]. Y. Bengio: From System 1 Deep Learning to System 2 Deep Learning | NeurIPS 2019
[3]. F Hutter, L Kotthoff, J Vanschoren. Automated machine learning: methods, systems, challenges. Book 2019

Outline

1. What is Machine Learning?
2. What is Automated Machine Learning (AutoML)?
   - Explanation from a Simple Example
   - Recent Industrial and Research Examples
   - A Generalization Viewpoint for AutoML
   - How to use AutoML
3. Summary & Next Works

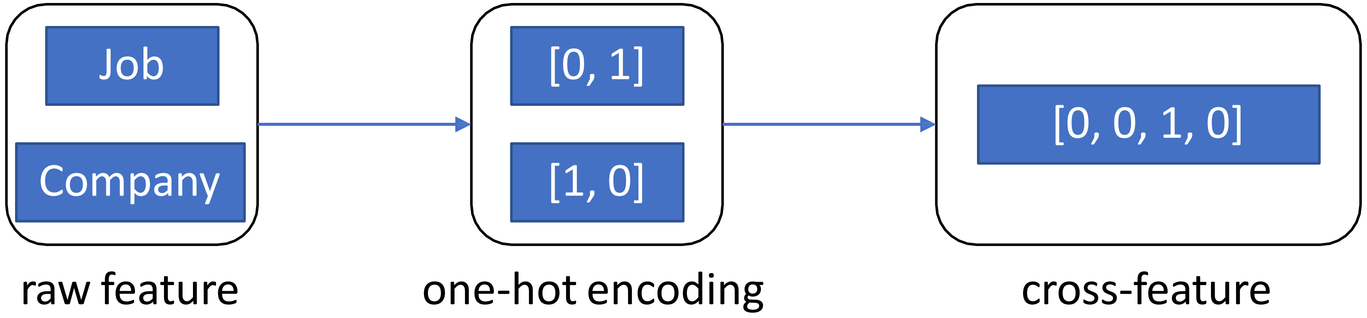
Industrial Example – Cross features

An example of tabular data (UCI-Bank)

	age (n)	job (c)	marital (c)	education (c)	balance (n)	housing (c)
0	30	unemployed	married	primary	1787	no
1	33	services	married	secondary	4789	yes
2	35	management	single	tertiary	1350	yes
3	30	management	married	tertiary	1476	yes
4	59	blue-collar	married	secondary	0	yes
5	35	management	single	tertiary	747	no

- Use one-hot/multi-hot encoding for categorical features
- Cross-features are empirically effective to enhance categorical features

Cross feature ‘job x company’ indicates that an individual takes a specific job in a specific company, and is a strong feature to predict one’s income



Not all cross-features are useful and too many of them lead to overfitting

How to find them?

[1]. Y. Luo, et.al. AutoCross: Automatic Feature Crossing for Tabular Data in Real-World Applications. KDD 2019

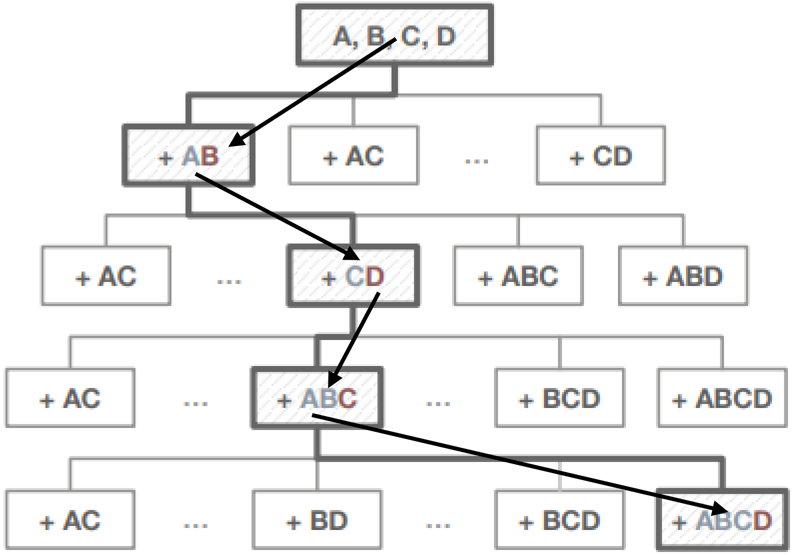
Industrial Example – AutoCross

Search cross features by bi-level optimization

$$\max_{\mathcal{S}\subseteq A(\mathcal{F})} \mathcal{E}\underbrace{(\overbrace{\mathcal{L}(\mathcal{D}_{tr},\mathcal{S})}^{1},\mathcal{D}_{vld},\mathcal{S})}_{2},$$

1. Obtain a classifier on training set with current cross-feature candidates
2. Measure cross-features' performance on validation set

All possible candidates



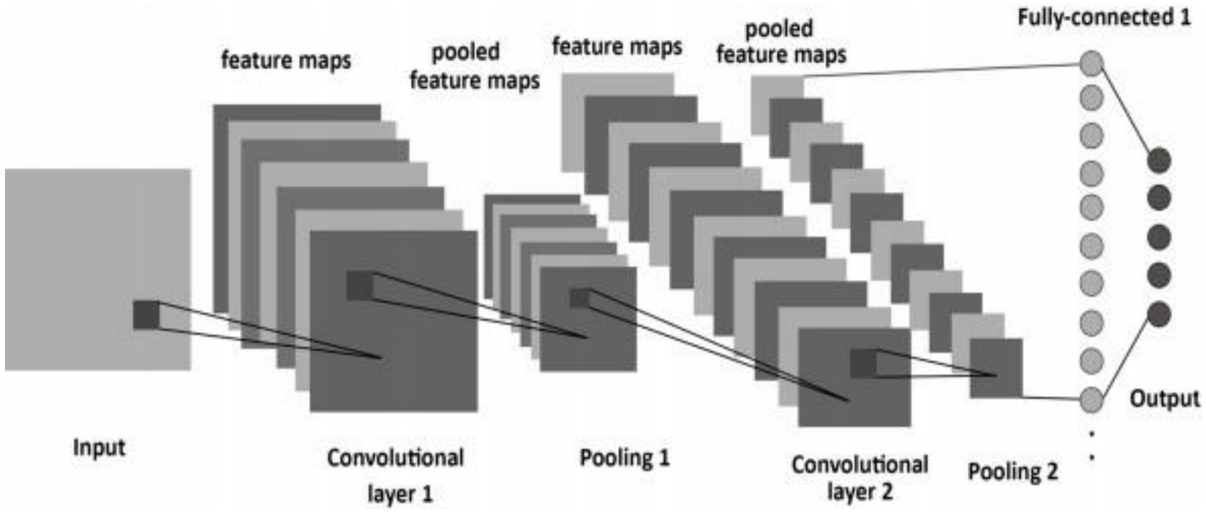
Candidate search process

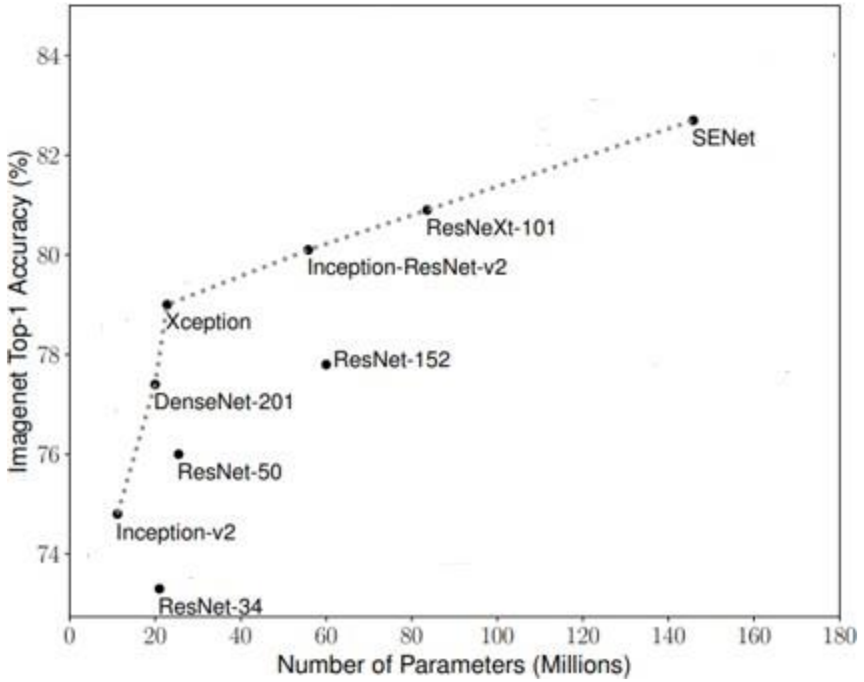
[1]. Y. Luo, et.al. AutoCross: Automatic Feature Crossing for Tabular Data in Real-World Applications. KDD 2019

Industrial Example – AutoCross

Video Link: https://www.youtube.com/watch?v=m47GcDR2Vak&t=9s

Academic Example – Neural archi. search (NAS)





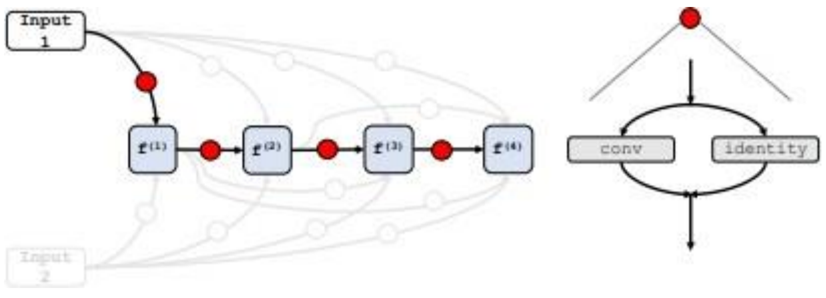
Design choice in each layer

- number of filters
- filter height
- filter width
- stride height
- stride width
- skip connections

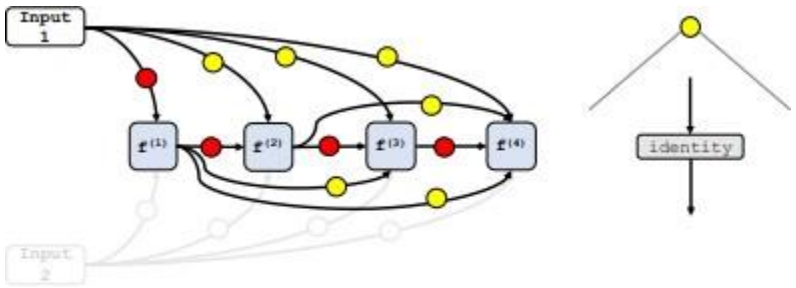
The design of architectures is important to CNN performance

Right figure: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. ICML 2019
Left figure: Gradient-based learning applied to document recognition. 1998

NAS – Search problem

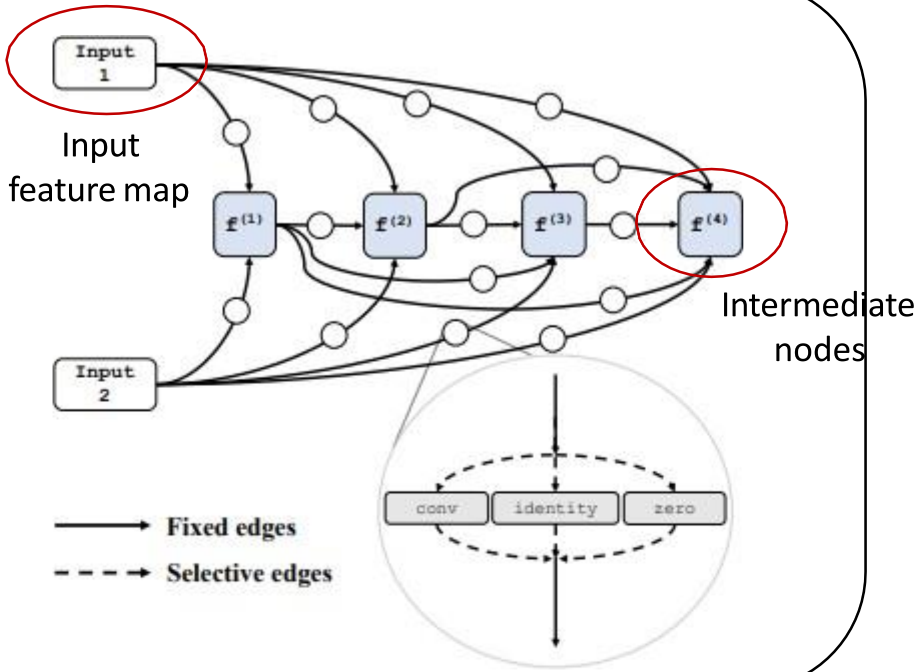


ResNet



DenseNet

Special cases

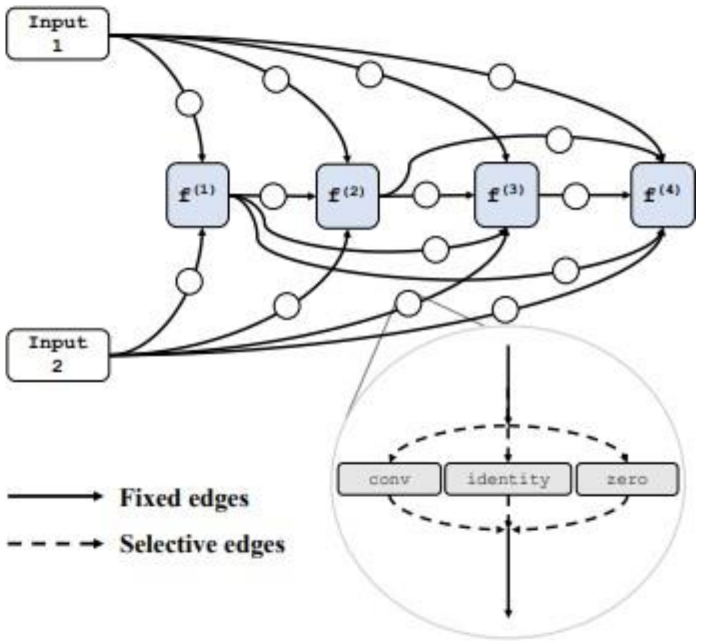


Example search space

- All possible candidates form a big supernet
- An architecture is a sub-graph in the supernet

Figures is from. M. Guo et.al. When NAS Meets Robustness: In Search of Robust Architectures against Adversarial Attacks. CVPR 2020

NAS – Search problem



Bi-level objective:

$$\min_{\mathbf{A}} \mathcal{F}(w^*, \mathbf{A}), \text{s.t.} \begin{cases} w^* = \arg\min_{w} \mathcal{L}_{\text{train}}(w, \mathbf{A}) \\ \mathbf{a}^{(i,j)} \in \mathcal{C} \end{cases}$$

- Train the selected architecture (encoding by $\mathbf{A}$) on training set
- Obtain the generalization performance of $\mathbf{A}$ on validation set

Typical search algorithms

- One-shot method[1,2] (fast but not accurate)
  - Alternative update architecture parameter $\mathbf{A}$ and network weights w^* by epochs
- Stand-alone method[3,4] (accurate but slow)
  - Obtain w^* by train network from scratch with given $\mathbf{A}$

[1]. H. Liu et al. Darts: Differentiable architecture search. ICLR 2018
[2]. A. Zela et al. Understanding and robustifying differentiable architecture search. ICLR 2020
[3]. Neural Architecture Search with Reinforcement Learning. ICLR 2017
[4]. K. Eggensperger. Et al. Efficient Benchmarking of Algorithm Configurators via Model-Based Surrogates. ML 2018

NAS – Promising performance

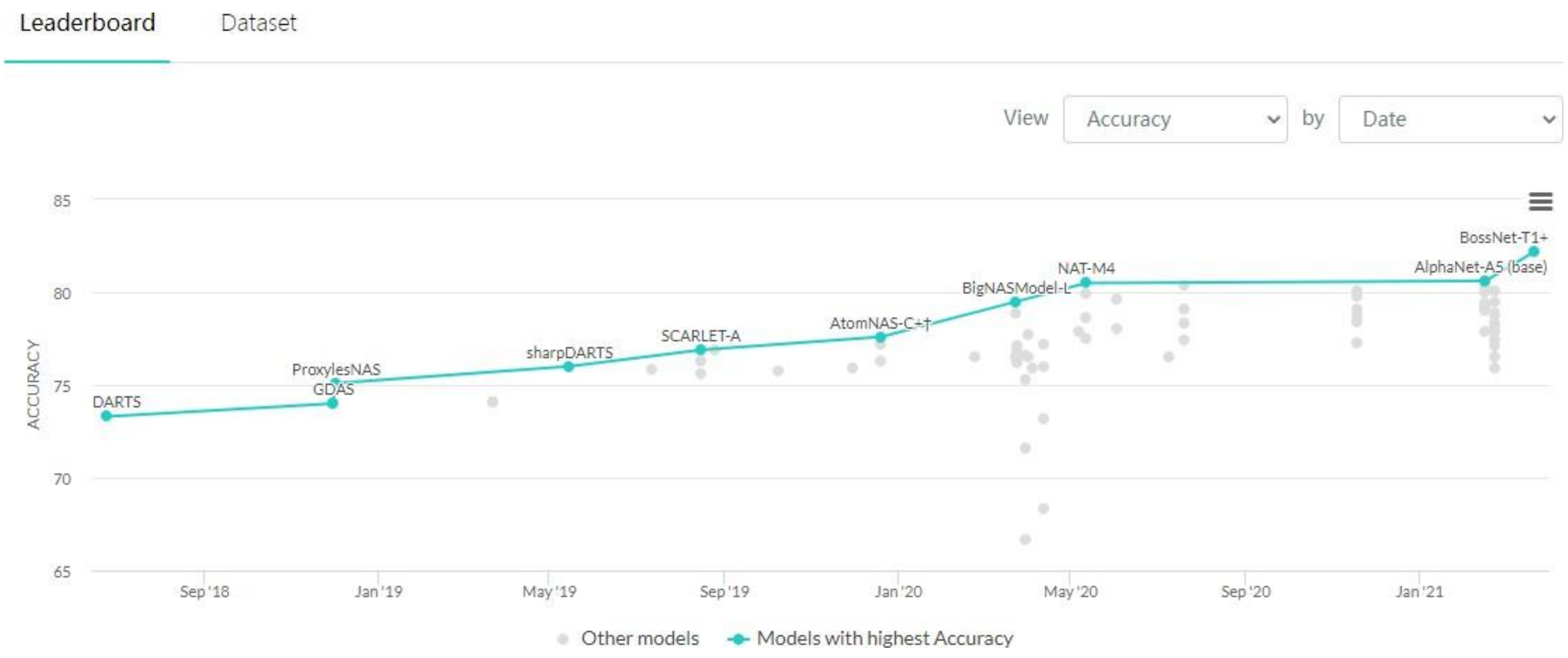


Figure is from: https://paperswithcode.com/sota/neural-architecture-search-on-imagenet

Outline

1. What is Machine Learning?
2. What is Automated Machine Learning (AutoML)?
    - Explanation from a Simple Example
    - Recent Industrial and Research Examples
    - A Generalization Viewpoint for AutoML
    - How to use AutoML
3. Summary & Next Works

Figure 4.1 & 4.2. M. Mohri, A. Rostamizadeh, A. Talwalkar. Foundations of machine learning. 2018
[1]. P. Battaglia. et.al. Relational inductive biases, deep learning, and graph networks. arXiv 2018.

What is AutoML – Generalization viewpoint

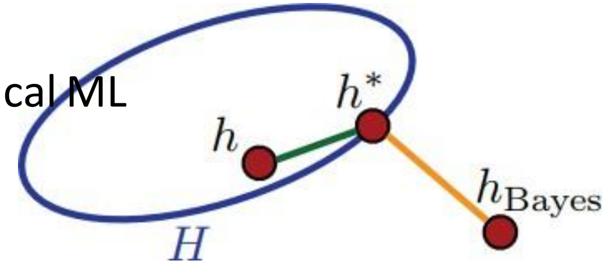
Parameterized the prior knowledge of learning methods, e.g.,

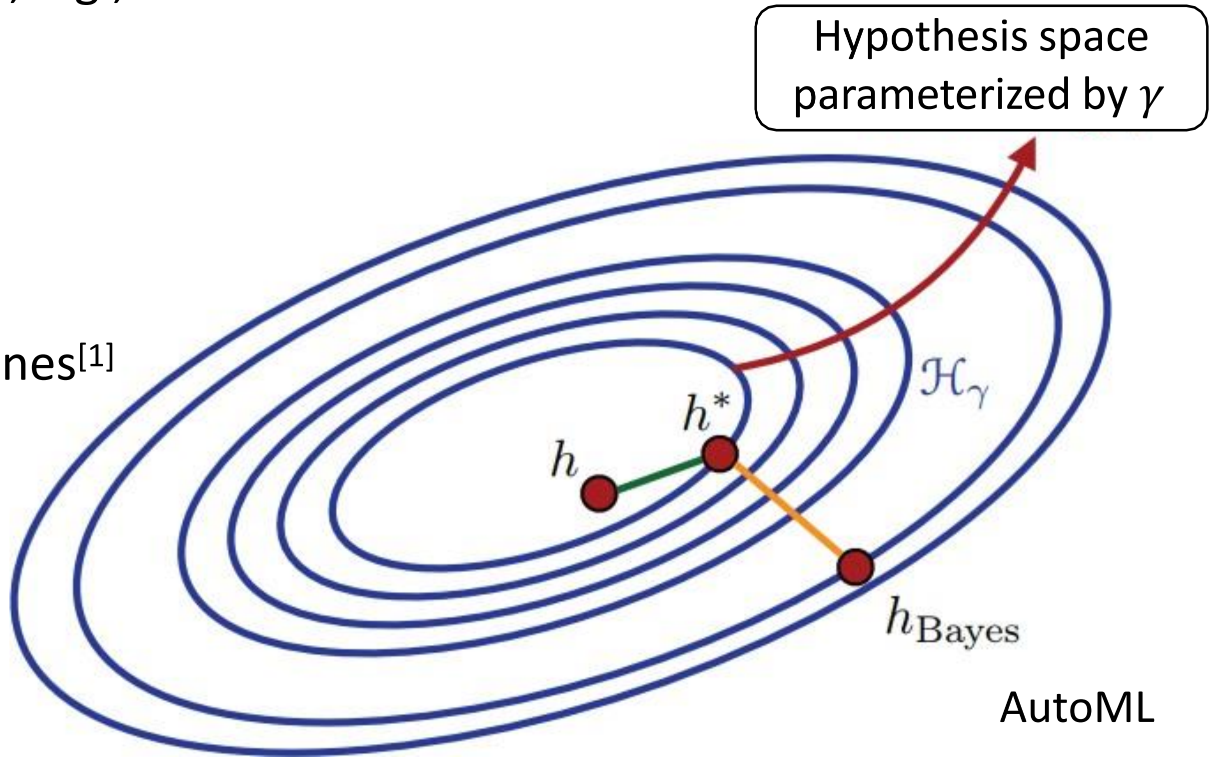
- minimize the total error
- reduce parameter numbers

Perform efficient search in the designed (new) space

- combinatorial generalize new models from existing ones[1]

Classical ML





AutoML

Parameterize (low-level) prior knowledge in the usage and design of machine learning

As a consequence

- Human participations can be naturally replaced by computation power
- total error of machine learning can be reduced (generalization can be improved)

AutoML – Successor of ML’s trend

- Core Issue in Machine Learning: Improving learning performance (with higher efficiency)
- AutoML: an evolving way to improve learning performance

Rule-based
1990s

Statistics-based
2000s

Deep Learning-based
2010s

AutoML-based
From 2017

Better performance

Continue the trends

- Larger hypothesis (more complex models) are being used
- Optimization is getting complex (even mixed up with generalization)
- The prior knowledge is imposed on more abstract level

Parameterize (low-level) prior knowledge in the usage and design of machine learning

Related Areas

Sub-areas

- Neural architecture search
- Hyper-parameter search
- Automated feature engineering
- Algorithms selection
- Model selection

Related areas

- Bi-level / Derivative-free optimization
  - Focus more on algorithm design
  - AutoML objective is one kind of objective where these algorithms can be applied
- Meta-learning
  - Focus on parameterize task distributions
  - Another kind of bi-level objective
  - Do not use validation set to update hyper-parameters

Outline

1. What is Machine Learning?
2. What is Automated Machine Learning (AutoML)?
   - Explanation from a Simple Example
   - Recent Industrial and Research Examples
   - A Generalization Viewpoint for AutoML
   - How to use AutoML
3. Summary & Next Works

How to use AutoML

1. Define an AutoML problem

- Derive a search space from insights in specific domains
- Search objective is usually validation performance
- Search constraint is usually resource budgets
- Training objective usually comes from classical learning models

Search Objective

Search Space

$$\min_{\lambda \in S} M\left(F(w^*;\lambda), D_{\text{val}}\right)$$

$$\text{s.t.} \begin{cases} \boxed{\min_{w} L\left(F(w;\lambda), D_{\text{tra}}\right)} \\ G(\lambda) \leq C \end{cases}$$

Training Objective

Search Constraints

2. Design or select proper search algorithm

- Reduce model training cost (time to get w^*)

What is AutoML – Short summary

- Exploring prior knowledge is important in machine learning
  - Cost time and critical to generalization performance
  - Continual trends in ML: imposing the prior knowledge on more abstract level
- AutoML attempts to parameterize low-level prior knowledge
  - Human participations can be naturally replaced by computation power
  - total error can be reduced (generalization can be improved)
- To use well AutoML techniques
  - Exploring high-level domain knowledge when defining the AutoML problem
  - Reducing model training cost when design search algorithm

Outline

1. What is Machine Learning?
2. What is Automated Machine Learning (AutoML)?
3. Summary & Next Works

Summary & Next Works

1. An introduction to Automated Machine Learning (AutoML)
   - Background on technical tools from machine learning
2. Automated Graph Neural Networks (GNN)
   - Explore neural architecture search for GNN
3. Automated Knowledge Graph (KG) Embedding
   - Explore AutoML for KG Embedding

Attach graph learning problems using AutoML

Works Covered

1. Pooling Architecture Search for Graph Classification. CIKM 2021
2. DiffMG: Differentiable Meta Graph Search for Heterogeneous Graph Neural Networks. KDD 2021
3. Efficient Relation-aware Scoring Function Search for Knowledge Graph Embedding. ICDE 2021
4. Search to aggregate neighborhood for graph neural network. ICDE 2021
5. Interstellar: Searching Recurrent Architecture for Knowledge Graph Embedding. NeurIPS 2020
6. AutoSF: Searching Scoring Functions for Knowledge Graph Embedding. ICDE 2020
7. AutoCross: Automatic Feature Crossing for Tabular Data in Real-World Applications. KDD 2019

Summary & Next Works

Research Topic

Automated Learning from Graph Structured Data

Manage topic | Submit your abstract | Submit your manuscript | Participate

Overview | Articles | Authors | Impact

About this Research Topic

Machine learning is an important technique to learn from graph structured data (GSD). However, since it is knowledge- and labor-intensive to pursue good learning performance, humans are heavily involved in every aspect of learning from GSD. Examples are property prediction for molecular graphs, product recommendations from heterogeneous information networks, and logical queries from knowledge base. To make machine learning techniques easier to apply and reduce the demand for experienced human experts, automated machine learning (AutoML) has emerged as a hot topic with great interest. At the same time, a special type of deep network model has been recently proposed, called graph neural networks (GNNs), which capture the dependence of graphs via message passing between the nodes of graphs and allow learning from GSD in a more principled manner. Moreover, they allow learning from molecular dynamics and meteorological simulation, which further extend the classical scope of GSD.

However, again due to such diversities in GSD, there is no unitary GSD learning model that can perform consistently across different tasks and datasets. As a consequence, how to design adaptive methods to learn from GSD in a task-aware and data-specific manner is an important problem. Fortunately, there are emerging technical tools from machine learning communities that have the potential to solve the above problems. Examples are automated machine learning, neural architecture search, meta-learning, and learning to lean. These methods can all help generalize learning methods that can exhibit abilities to learn well across different datasets and tasks. Thus, by proposing this Research Topic, we hope to draw interest from both academia and industry, with the goal to push learning methods for GSD to the next level.

Special Issue in: Frontiers in artificial intelligence

Jointly hold by: Quanming Yao, Huan Zhao

Rex Ying, and Xia Hu

Your submission is welcome!

Summary & Next Works

AH2: Automated Learning form Graph-Structured Data
Quanming Yao, Huan Zhao and Yongqi Zhang

Graph-structured data (GSD) is ubiquitous in real-life applications, which appears in many learning applications such as property prediction for molecular graphs, product recommendations from heterogeneous information networks, and logical queries from knowledge graphs. Recently, learning from graph-structured data has also become a research focus in the machine learning community. However, again due to such diversities in GSD, there are no universal learning models that can perform well and consistently across different learning applications based on graphs. In sharp contrast to this, convolutional neural networks work well on natural images, and transformers are good choices for text data. In this tutorial, we will talk about using automated machine learning (AutoML) as a tool to design learning models for GSD. Specifically, we will elaborate on what is AutoML, what kind of prior information from graphs can be explored by AutoML, and how insights can be generated from the searched models.

Quanming Yao

Tsinghua University

Dr. Quanming Yao is a tenure-track assistant professor at EE, Tsinghua University. He obtained his Ph.D. degree at the CSE of HKUST and was the founding leader of 4Paradigm INC's machine learning research team. He is a recipient of Forbes-30-Under-30 (China), Young Scientist Awards (HKIS), and Google Fellowship (machine learning).

Huan Zhao

4Paradigm

Dr. Huan Zhao is a senior researcher in 4Paradigm, leading the research on automated graph representation learning (AutoGraph) and real-world applications like retailing recommendation and bioinformatics. He has published various papers on top-tier venues like KDD, CIKM, AAAI, and TKDE. He obtained his Doctor Degree from HKUST in Jan. 2019.

Yongqi Zhang

4Paradigm

Dr. Yongqi Zhang is a senior researcher in 4Paradigm. He obtained the Ph.D. degree at CSE of HKUST. He has published many top-tier conference/journal papers as first-author in NeurIPS, ICDE, VLDB-J. His research interests focus on KG embedding and AutoML. He has been a reviewer for AAAI, IJCAI, CIKM and TKDE.

AAAI-22 Tutorial Forum

Thirty-Sixth Conference on Artificial Intelligence

February 23, 2022

Vancouver, BC, Canada

Our next stop

- more recent works will be included

Thanks!