Lecture 1: Cryptography and Quantum Computing

1936 Alan Turing's paper "On Computable Numbers …" laid the foundations of computing.

1976 Diffie & Hellman: Invented Public-key cryptography to provide security in e-commerce, securely exchange keys over a public channel. Birth of modern cryptography! (N, a, a^r, a^s)

1982 Yao MPC: solution for 2PC is garbled circuit protocol.

In 1925, Heisenberg invented matrix mechanics, Schrödinger invented wave mechanics, mark the birth of modern quantum mechanics.

n quantum bits are represented by a superposition state vector in 2^n dimensional Hilbert space.

1977 RSA (2002 图灵奖): $N = pq, ed = 1 \bmod \varphi(N)$, Alice 公开 e, Bob 公开 a^e

1994 Shor's Factoring Algorithm, 破解密码

Lecture 2: Computer Systems Crash Course

SRAM: 6 or 4 transistors, DRAM: 1 transistor + 1 capacitor CPU cache: On-chip SRAM

float: sign 1 + exp 8 + frac 23 $(-1)^{\text{sign}} \times (1.\text{frac})_2 \times 2^{\text{exp}_2 - 127}$

Von Neumann Architecture: Input → CPU (Control Unit + ALU) + Memory unit → Output

95 of the 128 ASCII code points are printable. Unicode encodes emoji as text.

CPU: Instruction Register → Decoding Unit, Program Counter, Clock, ALU

CPU cannot directly perform arithmetic operations through ALU for the data stored in memory.

Lecture 3: A Brief Overview of AI

Allen Newell(1975 图灵奖), Herbert Simon, and John McCarthy(1971 图灵奖) all joined Dartmouth workshop in 1956(the AI concept was introduced) Marvin Minsky (1969 图灵奖)

Noam Chomsky: invent generative grammar: describe & generate natural language, Probabilistic Context Free Grammar, logic + probabilistic inference → linguistics.

Lisp (1950): a favored programming language for AI research, self-modifying, one of the major data structures: Linked list, based on λ-calculus. (Large market with LISP machines)

First AI spring with logic programming: Edward Feigenbaum & Raj Reddy(1994 图灵奖)

Data-Driven methods: Judea Pearl(2011 图灵奖) Introduce probabilistic approaches and mathematical causality into ML, graphic models(Bayesian Networks)

The area of Bayesian machine learning: Unify statistics, human knowledge, ML.

The era of DL: Probabilistic Neural Language Model (1st NLM) → Seq2Seq (Google) → Transformer Model → Multi-Modal Language Model (GPT)

2018 图灵奖: • Geoffrey Hinton, early contributor of backpropagation and many fundamental ideas, the person that makes DL work

• Yann LeCun developed practical convolutional neural networks, foundation of modern pattern recognition. • Yoshua Bengio proposed neural language model and attention.

RL: Agent(policy), Environment and Reward, Learn by trial and error

Advanced human-AI interaction: diversity-driven RL, 挖掘可能的人类行为，根据玩家喜好调整

Python allows programmers to generate and execute new code during runtime using the eval function.

Minimax search: solve two-player zero-sum games, but too slow.

Alpha-beta prunning: exponentially time, heuristic algorithm (in contrast to exact algorithm)

Lecture 4: Introduction to Quantum Cryptography

Cryptographic Methods:

Symmetric: Same key for encryption and decryption, One-time pad (XOR, key use once)

Asymmetric: Mathematically related key pairs for encryption and decryption (RSA)

Kerckhoffs's Principle: The only secrecy involved with a cryptosystem should be the key.

Quantum cryptography doesn't rely on computational assumptions. It is distributing keys of the One-time pad algorithm.

QKD development: commercial and government application, expensive for personal

Copenhagen interpretation: one of the most commonly accepted (1920s, Bohr and Heisenberg)

Core of the debate: intrinsic randomness Unpredictable!

EPR Paradox: Entanglement(measure on one particle 立刻影响 the state of other), propose Local hidden variable (wrong, inconsistent with Bell test)

Bell test is a real-world physics experiment designed to test the theory of quantum mechanics in relation to Einstein's concept of local realism. based on Bell's theorem, Bell inequality violations can be used in quantum cryptography protocols to detect spies.

Non-Local game (CHSH): class limit $S_c = \frac{3}{4}, S_Q = \frac{2+\sqrt{2}}{4}$

EPR pair implies perfect (symmetric) key.

Take home messages: QKD systems distribute keys, doesn't replace all the current cryptosystems, doesn't replace current communication systems.

Lecture 5: A Decade of Machine Learning Accelerators (TPU and wrestle 哥)

ISA: define supported instructions, data types, registers, the hardware support for managing main memory, fundamental features, input/output model RISC: arm vs CISC: x86-64

David Patterson received his Turing Award (2017) for developing RISC, joined Google in 2016.

RISC facilitates the implementation of an instruction pipeline. CISC: 复杂指令，时间长

TPU: 自定义 Domain Specific Architecture 降低 DNN 推断成本 10×, good at handling workloads that require high-precision arithmetic

Energy limits modern chips, not number of transistors

External Memory access energy: 100× on chip memory access, 10000× arithmetic operation

TPU (an AI accelerator for NN): not suitable for LA programs that require frequent branching or contain many element-wise algebra operations. (并行 TPU > GPU > CPU)

XLA: a domain-specific compiler for LA, targets different hardware including C/GPU, takes models from PyTorch, TensorFlow, JAX etc. Single most important optimization: operator fusion (reduces memory needs)

KWh = Hours to train ✕ Number of Processors ✕ Average Power per Processor ✕ PUE(能源利用率)

4Ms for Energy Efficiency for ML: Model, Machine, Mechanization, Maps

Co-optimize 4Ms to realize the amazing potential of ML to positively impact many fields sustainably.

Lecture 6: Computer Science Theory

Early AI: Threshold Logic Unit(TLU, 阈值逻辑单元) 1943, McCulloch and Pitts

Stochastic gradient descent: Instead of taking the derivative of the sum, replace the sum by the sum of a small set of terms. It's cheaper than gradient descent in high-dimensional cases.

CNN: good at extracting local instead of global information, suitable for image classification.

SVM: most robust prediction methods, based on statistical learning frameworks & VC theory, perform non-linear classification using the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. used to perform hand-written character recognition.

Lecture 7: A Lecture on Learning Theory

$\text{Lte}(f)$ (test error) = $\min_{\hat{f} \in F} \text{Ltr}(\hat{f})$ (approximation error) + $\text{Ltr}(f) - \min_{\hat{f} \in F} \text{Ltr}(\hat{f})$ (optimization error) + $\text{Lte}(f) - \text{Ltr}(f)$(generalization error)

Universal Approximation Theorem: Intuitively, a perceptron network with a single infinitely wide hidden layer can approximate arbitrary functions, but doesn't guarantee generalizing to unseen data (DNN is still necessary) VC dim of fully connected NN (with W, H layers) is $\Theta(WH)$

NN (minimizing the cost function) is non-convex. Original: $x_{t+1} = x_t - \eta \nabla f(x_t)$

Momentum (Nesterov 1989): accelerate gradient descent $v_{t+1} = -\nabla f(x_t + \beta v_t) + \beta v_t, x_{t+1} = x_t + \eta v_{t+1}$

Random noise phenomenon (unexplained): NN fit random labels and true labels have same VCdim, optimization → 0 for both true labels and random labels.

Rademacher complexity: measure how well can a classifier class fit random noise.

NFL: all optimization methods perform equally well when averaging over all optimization tasks without re-sampling.

Lecture 8: Learning Structured World Models From and For Physical Interactions (Robotic)

Model-Based RL: (+) Learn from partial observation, handle unknown system.

(-) unstructured system, limited generalization
Analytical Physics-Based RL: (+) structured system, excellent generalization
(-) Require full state information, state & parameter estimation.
Learned model is naturally differentiable, apply MPC optimized using gradient descent (GPU)
Code-Generated 3D Value Maps (Vox Poser): LLM (convert human language into code) + VLM (detect things grounded in 3D robot perception), directly enables 6 DoF closed-loop actions, no robot data required.
Markov Decision Process (MDP) in RL isn't known in analytic form. A low discount factor γ motivates the decision maker to favor taking actions early, $\gamma \to 1$ 重视未来收益
Sim2real transfer: Skill-level planning + Motion-level planning, policy learning improves efficiency, more accurate than physics-based models (even with extensive system identification), isn't necessary step for building the world model.
Multi-modal sensing and learning platforms (touch + vision): Scalable Dense Tactile Glove
Domain randomization applies randomization to the environment during policy training. This helps the policy to generalize to the real world.

Lecture 9: Compositional Modeling of 3D Objects and Scenes (椅子哥)

ShapeNet: richly annotated(geometry, functional), large-scale, 3D CAD model, organizes models under synonym(同义词) sets according to WordNet
Hierarchical Graph Encoder → feature vector in shape space → Hierarchical Graph Decoder
perform interpolation(插值) in shape space to get new object.
Composition-Based Modeling: propose an iterative assembly system, suggest complementary parts and locations at each time.
Automatic Shape Synthesis: add the maximum probability part iteratively.
Autoencoder (encoder + decoder) is a NN to learn an identity function in an unsupervised way to reconstruct the original input while compressing the data to discover a more efficient and compressed representation. (高维输入，低维输出 feature vector)
Leonidas Guibas: name red-black tree, Guibas-Stolfi algorithm
3D reconstruction: active(主动打光获得深度) vs passive(测量不同角度, Stereo vision methods) method
Single view reconstruction: monocular(单眼) method, 不能完全重建

Lecture 10: Programming Languages: a crash landing

"Note G": TAC: Three Address Code, an important IR (Intermediate Representation) in compilers SSA: Static Single-Assignment, rename after each assign.
Turing Machine: infinite tape, alphabet Γ, states set Q, transitions δ between states and actions to take. (与λ-calculus 等价)
Church's λ-calculus: $(\lambda x.M)N = M[x:N]$, replace all x in M by N
C++, Java, Python support different paradigms, functional programming.
Two-parameter function: $(\lambda x.(\lambda y.F))AB = F[x:A][y:B]$
σ : mapping Var → Num environment, assign variable to value.
v: function $v(\text{Exp}, \sigma) \to$ Num, evaluates Exp under σ to value Num.
I: function $I(\text{Stmt}, \sigma) \to \sigma'$, interpret Stmt under σ, produce a new σ'
$v(\text{Num}, \sigma) \to \text{Num}\, v(\text{Var}, \sigma) \to \sigma(\text{Var})$ $I(\text{Var} := \text{Exp}, \sigma) = \sigma(\text{Var} : v\ (\text{Exp}, \sigma))$
$I(\text{Stmt1};\ \text{Stmt2}, \sigma) = I(\text{Stmt2};\ \text{I}(\text{Stmt1}, \sigma))$

Lecture 11: AI for Social Good

How: Descriptive, Predictive, Prescriptive(suggest) Major application domain: ML, health
Security(巡逻): Game model and Linear Programming-based solution –– Zero-sum, min Attacker's max expected utility, Flow-based Representation + Critical Time points
Poacher: Decision Trees + Markov Random Field (challenge: Lack of labeled data + data imbalance), 基于捕猎者模型 plan patrol
NewsPanda: Fine tune LM(BERT) to look through articles to identify trends, events of threats.
Transportation: Myopic surge cause regret, min cost flow using LP max social welfare (reject still exist), dual solution forms competitive equilibrium, driver still deviate to trigger recompute
Spatio-Temporal Pricing decrease time efficiency, regret → 0, social welfare++
Volunteer-based Food Rescue: stacking model(预测接单)
Optimize Intervention and Notification Scheme (INS) by Branch-and-Bound Algorithm.
Rescue-Specific Notification: 限定每人最多接到 L 条通知, 贪心选前 k, online planning

Lecture 12: Computing in Miniature: from Graphics to Science (bubble 哥)

Simulation: Geometric representation, to discretize thin volumes, films
common discretizations 导致大量粒子, design discrete analogs for fluid ––
Geometric Representation: Moving-Least-Squares (MLS) Particles / Moving Eulerian (内部 normal deformation)-Lagrangian (表面 tangential flow details) Particles (MELP)
Contact Dynamics (Solid-Fluid Interaction): 可变形网格 Tracking a Lagrangian Line
Design: Differentiable representations to characterize physics, geometry, and constraints
Discovery: Learning prior representations to embed math, physics, and numerical priors
Topology optimization: optimize the material layout in a space (shape optimization doesn't support arbitrary topology) $\rho(\partial \boldsymbol{u}/\partial t + \boldsymbol{u}\cdot\nabla\boldsymbol{u}) = -\nabla p + \mu\nabla^2\boldsymbol{u} + \boldsymbol{f}, \nabla\cdot\boldsymbol{u} = \boldsymbol{0}$
Computational challenges: curse of discretization dimensionality
Current Method: Voronoi drawn by hand, differentiable topological representation: differential Voronoi diagram. Build Physical Simulators without governing equations.

Lecture 13: ChatGLM: from LLM to AGI

符号 AI（知识可搜索）→ 感知智能（知识可计算）→ 认知智能（认知可计算）
静态表示：Word2Vec, doesn't require a large-scale human-annotated natural language dataset.
动态表示：Transformer Super Alignment: ensure safe to society.
BERT (自编码): masks some of the input tokes at random and then predicts those masked tokens, fine-tuned to model downstream tasks.
Autoregressive language model (GPT) predicts next work based on the word before
PageRank: first algorithm to order search results by Google, assigns a numerical weighting to each element of a hyperlinked set of documents (WWW)
• continuous bag-of-words (CBOW): semantically similar words should be in similar contexts, 根据上下文预测中间, faster and better accuracy for the frequent words
• continuous skip-gram architecture, 根据中间预测上下文, better at small amount of the training data and rare words

Lecture 14: 面向下一代人工智能的高能效电路与系统设计

能效提升：Scaling down，定制加速器，新器件+新模型（高能效存内计算新架构）
软硬件协同优化：面向神经网络，容错度较高, enable analog circuits in accelerator.
应用场景：云端(服务器，计算+存储)，边缘(基站，路由器，计算+通信)，终端(显示+交互)
存算一体架构：RRAM 等忆阻器,$O(1)$矩阵向量乘法, 无需搬数据(冯诺伊曼架构主要能耗在搬数据)
AI 1.0: 用于判别任务的专用模型 AI 2.0: 生成任务的通用模型
算法创新优化(10 - 30) + 模型计算优化(10 -20) + 算力平台优化(5 - 10) + 硬件推理优化(10 - 20)
CPU which strictly follows the von Neumann architecture only perform 1 instruction at one time.
GPU: 并行加速 NPU: a class of specialized hardware accelerators or computer systems designed to accelerate AI applications.