So Far: Foundational Methods
Now: Advanced Applications

You will see later

Después lo veras

[Images of advanced applications: a helicopter, a robot, and a vehicle with sensors and equipment]
Google is trying to make artificial intelligence history — and it could happen this week

Drake Baer  
Mar. 7, 2016, 3:49 PM  
9,639
In Two Moves, AlphaGo and Lee Sedol Redefined the Future

Seoul, South Korea — In Game Two, the Google machine made a move that no human ever would. And it was beautiful. As the world looked on, the move so perfectly demonstrated the enormously powerful and rather
How would you make an AI for Go?
MiniMax!
Why is it hard?

In particular, why is it harder than chess?
Exhaustive search
Reducing depth with value network
Reducing breadth with policy network
Neural network training pipeline

- Human expert positions
  - Supervised Learning policy network
  - Reinforcement Learning policy network
  - Self-play data
  - Value network

**N-Layer Neural Network**

**Policy Search**
- **Simplest policy search:**
  - Start with an initial linear value function or Q-function
  - Nudge each feature weight up and down and see if your policy is better than before

- **Problems:**
  - How do we tell the policy got better?
  - Need to run many sample episodes!
  - If there are a lot of features, this can be impractical

- **Better methods exploit lookahead structure, sample wisely, change multiple parameters...**
One more thing: Monte-Carlo rollouts
Mastering the game of Go with deep neural networks and tree search

David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel & Demis Hassabis

Affiliations | Contributions | Corresponding authors

Nature 529, 484–489 (28 January 2016) | doi:10.1038/nature19661
Received 11 November 2015 | Accepted 05 January 2016 | Published online 27 January 2016
Mastering the game of Go without human knowledge


Artificial intelligence: Learning to play Go from scratch
Satinder Singh, Andy Okun & Andrew Jackson
Robotic Helicopters
Motivating Example

- How do we execute a task like this?

[VIDEO: tictoc_results.wmv]
Key challenges:

- Track helicopter position and orientation during flight
- Decide on control inputs to send to helicopter
Autonomous Helicopter Setup

On-board inertial measurement unit (IMU)

Send out controls to helicopter

Position
HMM for Tracking the Helicopter

- **State:** \( s = (x, y, z, \phi, \theta, \psi, \dot{x}, \dot{y}, \dot{z}, \dot{\phi}, \dot{\theta}, \dot{\psi}) \)
- **Measurements:** [observation update]
  - 3-D coordinates from vision, 3-axis magnetometer, 3-axis gyro, 3-axis accelerometer
- **Transitions (dynamics):** [time elapse update]
  - \( s_{t+1} = f(s_t, a_t) + w_t \)  
    - \( f \) encodes helicopter dynamics, \( w \): noise
Helicopter MDP

- **State:** \( s = (x, y, z, \phi, \theta, \psi, \dot{x}, \dot{y}, \dot{z}, \dot{\phi}, \dot{\theta}, \dot{\psi}) \)

- **Actions (control inputs):**
  - \( a_{\text{lon}} \): Main rotor longitudinal cyclic pitch control (affects pitch rate)
  - \( a_{\text{lat}} \): Main rotor latitudinal cyclic pitch control (affects roll rate)
  - \( a_{\text{coll}} \): Main rotor collective pitch (affects main rotor thrust)
  - \( a_{\text{rud}} \): Tail rotor collective pitch (affects tail rotor thrust)

- **Transitions (dynamics):**
  - \( s_{t+1} = f(s_t, a_t) + w_t \)
    - \( f \) encodes helicopter dynamics
    - \( w \) is a probabilistic noise model

- **Can we solve the MDP yet?**
Problem: What’s the Reward?

- Reward for hovering:

\[
R(s) = -\alpha_x (x - x^*)^2 \\
-\alpha_y (y - y^*)^2 \\
-\alpha_z (z - z^*)^2 \\
-\alpha_x \ddot{x}^2 \\
-\alpha_y \dot{y}^2 \\
-\alpha_z \dot{z}^2
\]
RL: Helicopter Flight
Problem for More General Case: What’s the Reward?

- Rewards for “Flip”?  
  - Problem: what’s the target trajectory?  
  - Just write it down by hand?
Flips (?)

[VIDEO: 20061204---bad.wmv]
Helicopter Apprenticeship?
Demonstrations

[VIDEO: airshow_unaligned.wmv]
Learning a Trajectory

- HMM-like generative model
  - Dynamics model used as HMM transition model
  - Demos are observations of hidden trajectory
- Problem: how do we align observations to hidden trajectory?
Probabilistic Alignment using a Bayes’ Net

- Dynamic Time Warping
- Extended Kalman filter / smoother

Abbeel, Coates, Ng, IJRR 2010
Aligned Demonstrations

[VIDEO: airshow_unaligned.wmv]
Alignment of Samples

- Result: inferred sequence is much cleaner!
Learned Behavior

[VIDEO: airshow_trimmed.wmv]

[Abbeel, Coates, Quigley, Ng, 2010]
Legged Locomotion
For Perspective: Darpa Robotics Challenge (2015)
How About Continuous Control, e.g., Locomotion?

Robot models in physics simulator (MuJoCo, from Emo Todorov)

Input: joint angles and velocities
Output: joint torques

Neural network architecture:
Learning Locomotion

[Schulman, Moritz, Levine, Jordan, Abbeel, 2015]
Deep RL: Virtual Stuntman

[Peng, Abbeel, Levine, van de Panne, 2018]
Quadruped

- Low-level control problem: moving a foot into a new location → search with successor function ~ moving the motors

- High-level control problem: where should we place the feet?
  - Reward function $R(x) = w \cdot f(s)$ [25 features]

[Kolter, Abbeel & Ng, 2008]
Reward Learning + Reinforcement Learning

- Demonstrate path across the “training terrain”

- Learn the reward function

- Receive “testing terrain”---height map.

- Find the optimal policy with respect to the learned reward function for crossing the testing terrain.

[Kolter, Abbeel & Ng, 2008]
Without reward learning
With reward learning
Autonomous Driving
Grand Challenge 2005: Barstow, CA, to Primm, NV

- 150 mile off-road robot race across the Mojave desert
- Natural and manmade hazards
- No driver, no remote control
- No dynamic passing
Autonomous Vehicles

Autonomous vehicle slides adapted from Sebastian Thrun
Grand Challenge 2005 Nova Video

[VIDEO: nova-race-supershort.mp4]
Grand Challenge 2005 – Bad
An Autonomous Car

- Lasers
- Camera
- Radar
- GPS
- GPS compass
- 6 Computers
- IMU
- E-stop
- Control Screen
- Steering motor
Actions: Steering Control

- Reference Trajectory
- Error
- Velocity
- Steering Angle (with respect to trajectory)
Laser Readings for Flat / Empty Road
Laser Readings for Road with Obstacle
Obstacle Detection

Trigger if $|Z_i - Z_j| > 15\text{cm}$ for nearby $z_i, z_j$

Raw Measurements: 12.6% false positives
Probabilistic Error Model
HMMs for Detection

Raw Measurements: 12.6% false positives

HMM Inference: 0.02% false positives
Sensors: Camera
Vision for a Car
Vision for a Car

[VIDEO: lidar vision for a car]
Self-Supervised Vision

[VIDEO: self-supervised vision]
Urban Environments
Google Self-Driving Car (2013)

(mostly lidar)
Recent Progress: NN Semantic Scene Segmentation

PSPNet50

~ neural net classifies every pixel
Self-Driving Cars -- Stats

Autonomous vehicle safety progress

Events/10,000 miles

Google/Waymo disengagements per 1000 miles
Cruise/GM disengagements per 1000 miles
Nissan disengagements per 1000 miles
Delphi disengagements per 1000 miles
Human crash rate per 1000 miles (0.002-0.004)
Human injuries per 1000 miles (0.00077)
Human fatalities per 1000 miles (0.00003)

2014 2015 2016 2017

Years

Autonomous vehicle safety progress (log scale)

Events/10,000 miles

Google/Waymo disengagements per 1000 miles
Cruise/GM disengagements per 1000 miles
Nissan disengagements per 1000 miles
Delphi disengagements per 1000 miles
Human crash rate per 1000 miles (0.002-0.004)
Human injuries per 1000 miles (0.00077)
Human fatalities per 1000 miles (0.00003)

2014 2015 2016 2017

Years
Energy-Inference-Accuracy Landscape on the Squeezelator

SqueezeNext vs SqueezeNet/AlexNet

- 8% more accurate
- 2.25x better than SqueezeNet
- 7.5x better than AlexNet

* MobileNet v1

[slide credit: Kurt Keutzer]
Personal Robotics
Challenge Task: Robotic Laundry
Sock Sorting

Five previously unseen socks are placed on the table.
How about a range of skills?
Learned Skills
Unsupervised Learning for Interaction?

[Levine et al, 2016]
Next Time:

- Natural language processing
- Final contest
- Course wrap-up
- Where to go next