

The State of Techniques for Solving Large Imperfect-Information Games

Tuomas Sandholm

Professor

Carnegie Mellon University

Computer Science Department

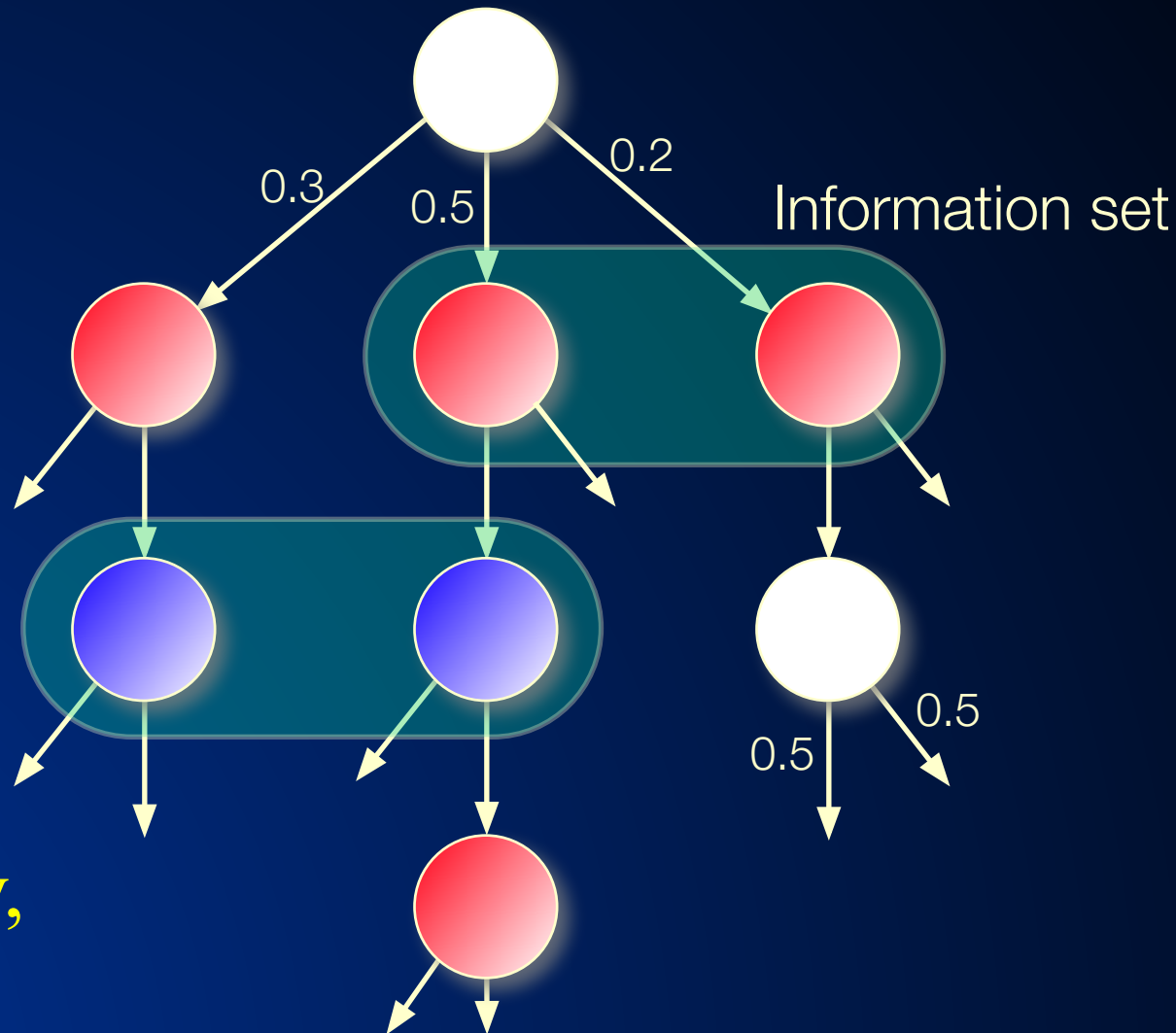
Also:

Machine Learning Department

Ph.D. Program in Algorithms, Combinatorics, and Optimization

CMU/UPitt Joint Ph.D. Program in Computational Biology

Incomplete-information game tree



Tackling such games

- Domain-independent techniques
- Techniques for complete-info games don't apply
- Challenges
 - Unknown state
 - Uncertainty about what other agents and nature will do
 - Interpreting signals and avoiding signaling too much
- **Definition.** A **Nash equilibrium** is a *strategy* and *beliefs* for each agent such that no agent benefits from using a different strategy
 - Beliefs derived from strategies using Bayes' rule

Most real-world games are like this

- Negotiation
- Multi-stage auctions (FCC ascending, combinatorial)
- Sequential auctions of multiple items
- Political campaigns (TV spending)
- Military (allocating troops; spending on space vs ocean)
- Next-generation (cyber)security (jamming [DeBruhl et al.]; OS)
- Medical treatment [Sandholm 2012, AAI-15 SMT Blue Skies]
- ...

Poker

Recognized challenge problem in AI since 1992 [Billings, Schaeffer, ...]

- Hidden information (other players' cards)
- Uncertainty about future events
- Deceptive strategies needed in a good player
- Very large game trees

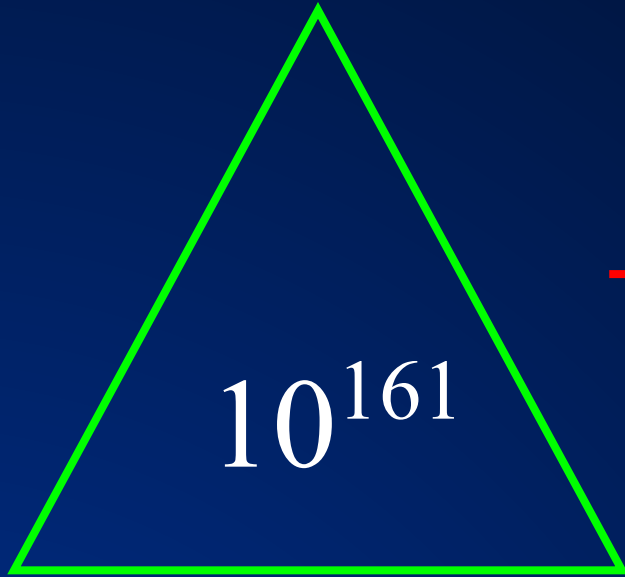
NBC National Heads-Up Poker Championship 2013



Our approach [Gilpin & Sandholm EC-06, *J. of the ACM* 2007...]

Now used basically by all competitive Texas Hold'em programs

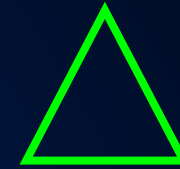
Original game



Automated abstraction



Abstracted game



Custom
equilibrium-finding
algorithm



Nash equilibrium

Reverse model



Nash equilibrium

Lossless abstraction

Information filters

- **Observation:** We can make games smaller by filtering the information a player receives
- Instead of observing a specific signal exactly, a player instead observes a **filtered set** of signals
 - *E.g.* receiving signal $\{A\spadesuit, A\clubsuit, A\heartsuit, A\diamondsuit\}$ instead of $A\heartsuit$

Solved Rhode Island Hold'em poker

- AI challenge problem [Shi & Littman 01]
 - 3.1 billion nodes in game tree
- Without abstraction, LP has 91,224,226 rows and columns => unsolvable
- *GameShrink* ran in one second
- After that, LP had 1,237,238 rows and columns (50,428,638 non-zeros)
- Solved the LP
 - CPLEX *barrier* method took 8 days & 25 GB RAM
- **Exact** Nash equilibrium
- **Largest incomplete-info game solved by then by over 4 orders of magnitude**



Lossy abstraction

Texas Hold'em poker



- 2-player Limit has $\sim 10^{14}$ info sets
- 2-player No-Limit has $\sim 10^{161}$ info sets
- Losslessly abstracted game too big to solve
=> abstract more
=> lossy

Important ideas for practical game abstraction 2007-13

- Integer programming [Gilpin & Sandholm AAMAS-07]
- Potential-aware [Gilpin, Sandholm & Sørensen AAI-07, Gilpin & Sandholm AAI-08]
- Imperfect recall [Waugh et al. SARA-09, Johanson et al. AAMAS-13]

Leading practical abstraction algorithm: Potential-aware imperfect-recall abstraction with earth-mover's distance

[Ganzfried & Sandholm AAI-14]

- Bottom-up pass of the tree, clustering using histograms over next-round clusters
 - EMD is now in multi-dimensional space
 - Ground distance assumed to be the (next-round) EMD between the corresponding cluster means

Techniques used to develop *Tartanian7*, program that won the heads-up no-limit Texas Hold'em ACPC-14

[Brown, Ganzfried, Sandholm AAMAS-15]

- Enables massive distribution or leveraging ccNUMA
- Abstraction:
 - **Top** of game abstracted with any algorithm
 - **Rest** of game split into equal-sized disjoint pieces based on public signals
 - This (5-card) abstraction determined based on transitions to a *base abstraction*
 - At each later stage, abstraction done within each piece separately
- Equilibrium finding (see also [Jackson, 2013; Johanson, 2007])
 - “Head” blade handles **top** in each iteration of External-Sampling MCCFR
 - Whenever the **rest** is reached, sample (a flop) from each public cluster
 - Continue the iteration on a separate blade for each public cluster. Return results to head node
 - Details:
 - Must weigh each cluster by probability it would've been sampled randomly
 - Can sample multiple flops from a cluster to reduce communication overhead



Lossy Game Abstraction with Bounds

Lossy game abstraction with bounds

- Tricky due to abstraction pathology [Waugh et al. AAMAS-09]
- Prior lossy abstraction algorithms had no bounds
 - First exception was for stochastic games only [S. & Singh EC-12]
- We do this for general extensive-form games [Kroer & S. EC-14]
 - Many new techniques required
 - For both action and state abstraction
 - More general abstraction operations by also allowing one-to-many mapping of nodes

Bounding abstraction quality

Main theorem:

For any Nash equilibrium σ' in M' , any undivided lifted strategy σ is an ϵ -Nash equilibrium in M

where $\epsilon = \max_{i \in \text{Players}} \epsilon_i$

$$\epsilon_i = 2\epsilon_i^R + \sum_{j \in \mathcal{H}_i} \epsilon_j^0 \overline{W} + \sum_{j \in \mathcal{H}_0} 2\epsilon_j^0 \overline{W}$$

Diagram illustrating the components of the error bound ϵ_i :

- Reward error**: $2\epsilon_i^R$
- Nature distribution error at height j**: $\epsilon_j^0 \overline{W}$ (for $j \in \mathcal{H}_i$)
- Maximum utility in abstract game**: \overline{W}
- Set of heights for player i**: \mathcal{H}_i
- Set of heights for nature**: \mathcal{H}_0
- Nature distribution error at height j**: $2\epsilon_j^0 \overline{W}$ (for $j \in \mathcal{H}_0$)

Hardness results

- Determining whether two subtrees are “extensive-form game-tree isomorphic” is graph isomorphism complete
- Computing the minimum-size abstraction given a bound is NP-complete
 - Holds also for minimizing a bound given a maximum size
- Doesn't mean abstraction with bounds is undoable or not worth it computationally

Extension to imperfect recall

- Merge information sets
- Allows payoff error
- Allows chance error

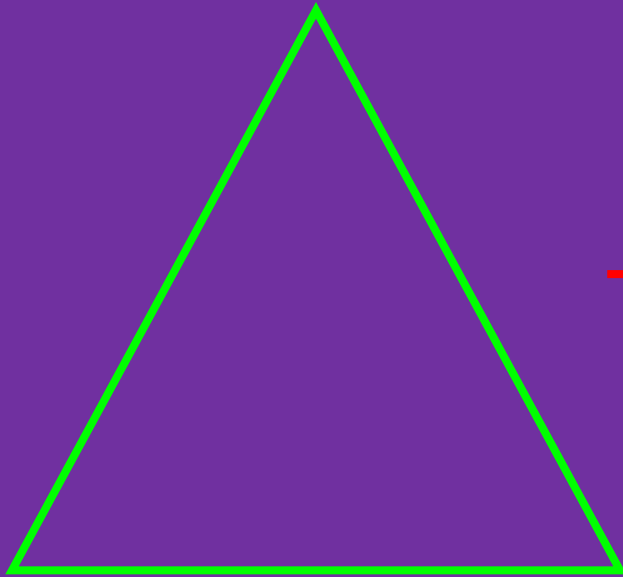
- Going to imperfect-recall setting costs an error increase that is linear in game-tree height

- Exponentially stronger bounds and broader class (abstraction can introduce nature error) than [Lanctot et al. ICML-12], which was also just for CFR

Role in modeling

- All modeling is abstraction
- These are the first results that tie game modeling choices to solution quality in the actual world!

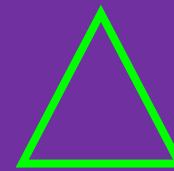
Original game



Automated abstraction



Abstracted game



Custom
equilibrium-finding
algorithm



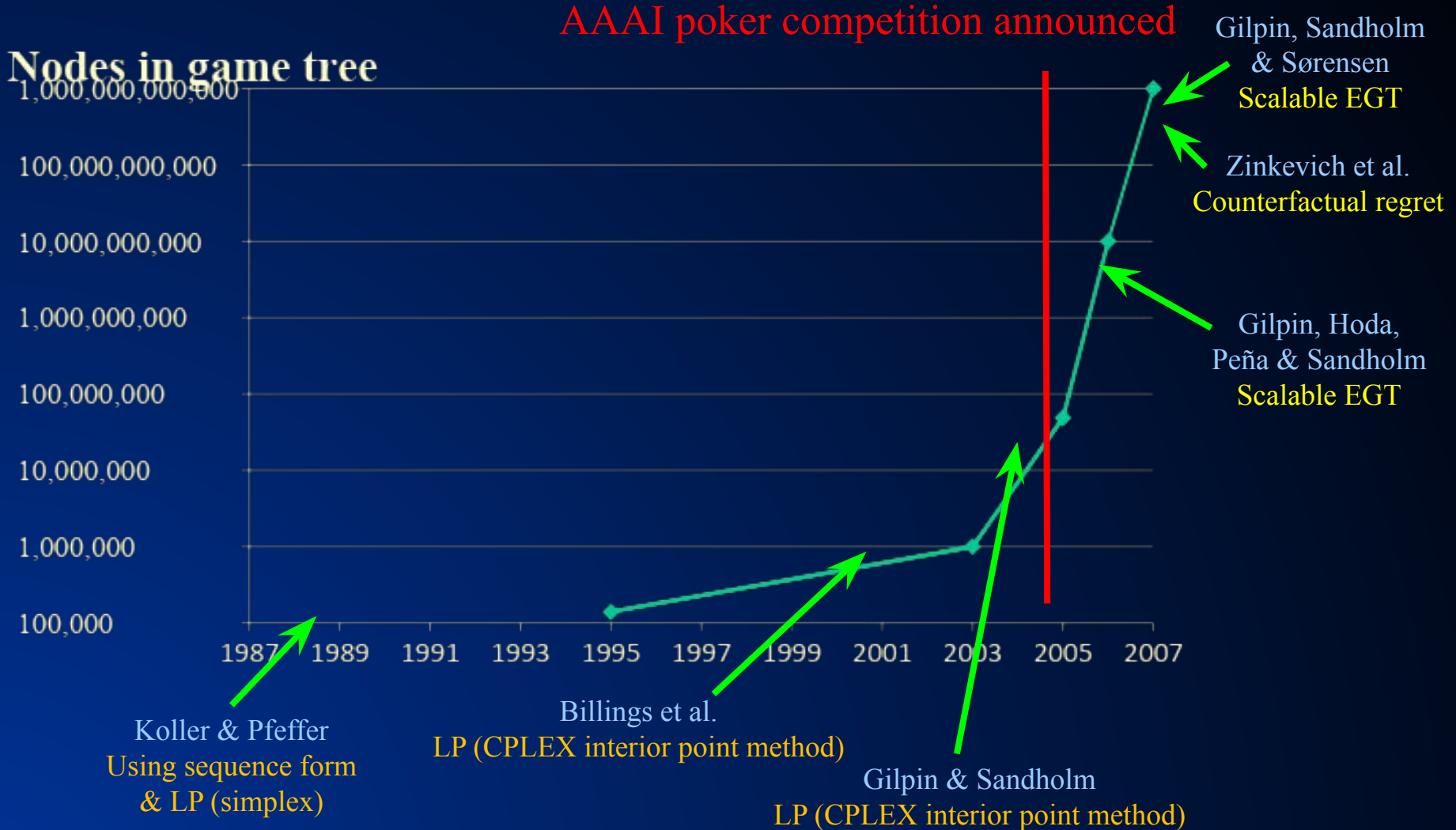
Nash equilibrium

Reverse model



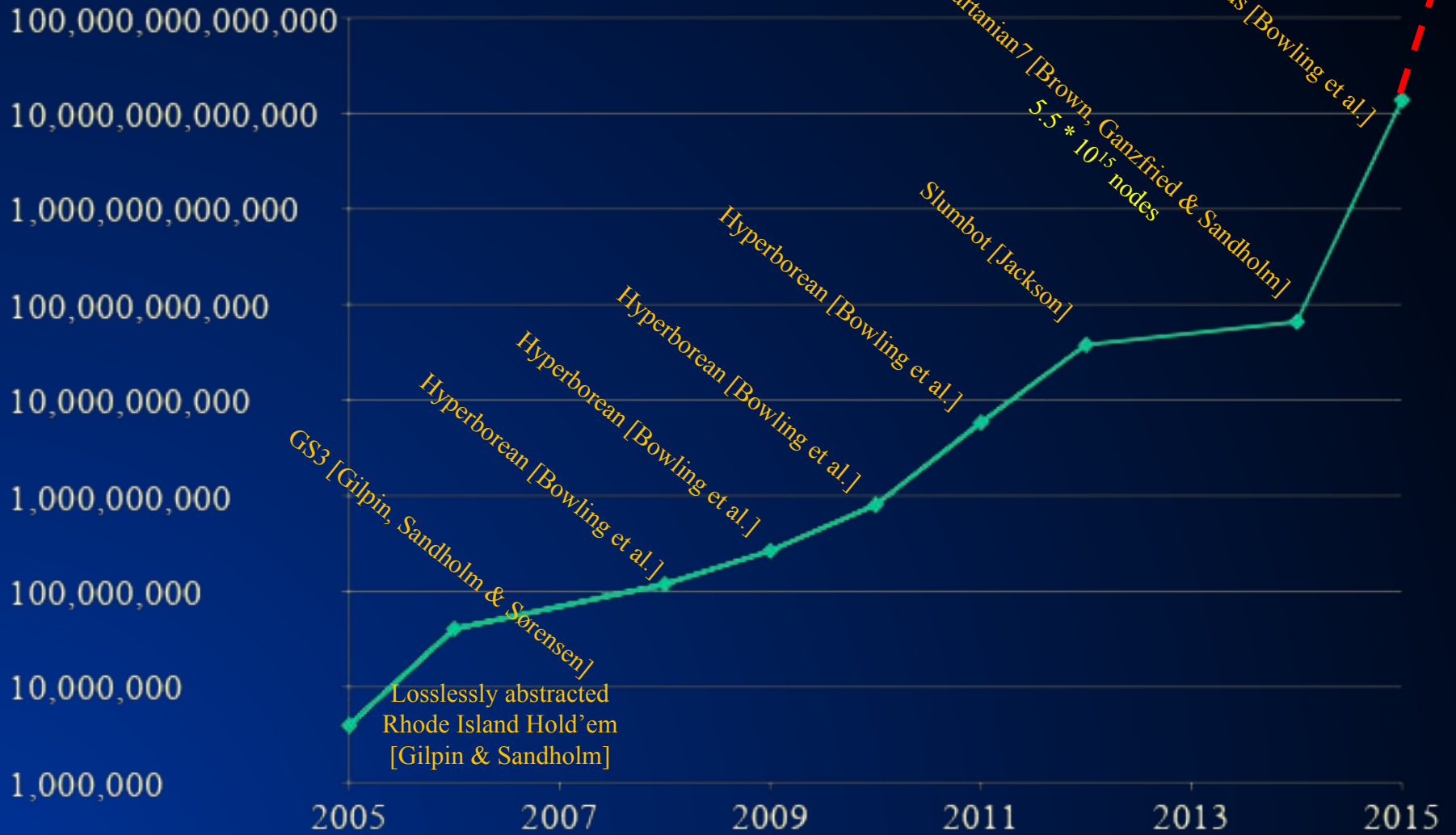
Nash equilibrium

Scalability of (near-)equilibrium finding in 2-player 0-sum games



Scalability of (near-)equilibrium finding in 2-player 0-sum games...

Information sets



Leading equilibrium-finding algorithms for 2-player 0-sum games

Counterfactual regret (CFR)

- Based on no-regret learning
- Most powerful innovations:
 - Each information set has a separate no-regret learner [Zinkevich et al. NIPS-07]
 - Sampling [Lanctot et al. NIPS-09, ...]
- $O(1/\epsilon^2)$ iterations
 - Each iteration is fast
- Parallelizes
- Selective superiority
- Can be run on imperfect-recall games and with >2 players (without guarantee of converging to equilibrium)

Scalable EGT

- Based on Nesterov's Excessive Gap Technique
- Most powerful innovations:
 - [Hoda, Gilpin, Peña & Sandholm WINE-07, *Mathematics of Operations Research* 2011]
 - Smoothing fns for sequential games
 - Aggressive decrease of smoothing
 - Balanced smoothing
 - Available actions don't depend on chance => memory scalability
- $O(1/\epsilon)$ iterations
 - Each iteration is slow
- Parallelizes
- New $O(\log(1/\epsilon))$ algorithm [Gilpin, Peña & Sandholm AAAI-08, *Mathematical Programming* 2012]

Better first-order methods

[Kroer, Waugh, Kılınç-Karzan & Sandholm EC-15]

- New prox function for first-order methods such as EGT and Mirror Prox
 - Gives first explicit convergence-rate bounds for general zero-sum extensive-form games (prior explicit bounds were for very restricted class)
 - In addition to generalizing, bound improvement leads to a linear (in the worst case, quadratic for most games) improvement in the dependence on game specific constants
- Introduces gradient sampling scheme
 - Enables the first stochastic first-order approach with convergence guarantees for extensive-form games
 - As in CFR, can now represent game as tree that can be sampled
- Introduces first first-order method for imperfect-recall abstractions
 - As with other imperfect-recall approaches, not guaranteed to converge

Computing equilibria by leveraging qualitative models

Player 1's strategy Player 2's strategy

Weaker hand
↑
↓
Stronger hand

BLUFF-FOLD	FOLD/BLUFF
CHECK-FOLD	FOLD/CHECK
	BLUFF/CHECK
CHECK-CALL	CALL/CHECK
BET-FOLD	CALL/BET
	RAISE/BET

BLUFF-FOLD	FOLD/BLUFF
CHECK-FOLD	FOLD/CHECK
	BLUFF/CHECK
CHECK-CALL	CALL/CHECK
BET-FOLD	CALL/BET
	RAISE/BET

	BLUFF
CHECK-FOLD	
	CHECK
CHECK-CALL	
	BET

- **Theorem.** Given F_1, F_2 , and a qualitative model, we have a complete mixed-integer linear feasibility program for finding an equilibrium
- Qualitative models can enable proving existence of equilibrium & solve games for which algorithms didn't exist

[Ganzfried & Sandholm AAMAS-10 & newer draft]

Simultaneous Abstraction and Equilibrium Finding in Games

[Brown & Sandholm IJCAI-15 & new manuscript]

Problems solved

- Cannot solve without abstracting, and cannot principally abstract without solving
 - SAEF abstracts and solves simultaneously
- Must restart equilibrium finding when abstraction changes
 - SAEF does not need to restart (uses discounting)
- Abstraction size must be tuned to available runtime
 - In SAEF, abstraction increases in size over time
- Larger abstractions may not lead to better strategies
 - SAEF guarantees convergence to a full-game equilibrium

OPPONENT EXPLOITATION

Traditionally two approaches

- **Game theory approach** (abstraction+equilibrium finding)
 - Safe in 2-person 0-sum games
 - Doesn't maximally exploit weaknesses in opponent(s)
- **Opponent modeling**
 - Needs prohibitively many repetitions to learn in large games (loses too much during learning)
 - Crushed by game theory approach in Texas Hold'em
 - Same would be true of no-regret learning algorithms
 - *Get-taught-and-exploited problem* [Sandholm AIJ-07]

Let's hybridize the two approaches

- Start playing based on pre-computed (near-)equilibrium
- As we learn opponent(s) deviate from equilibrium, adjust our strategy to exploit their weaknesses
 - Adjust more in points of game where more data now available
 - Requires no prior knowledge about opponent
- Significantly outperforms game-theory-based base strategy in 2-player limit Texas Hold'em against
 - trivial opponents
 - weak opponents from AAAI computer poker competitions
- Don't have to turn this on against strong opponents

Other modern approaches to opponent exploitation

- ϵ -safe best response

[Johanson, Zinkevich & Bowling NIPS-07, Johanson & Bowling AISTATS-09]



- Precompute a small number of strong strategies.
Use no-regret learning to choose among them

[Bard, Johanson, Burch & Bowling AAMAS-13]

Safe opponent exploitation

- Definition. *Safe* strategy achieves at least the value of the (repeated) game in expectation
- Is safe exploitation possible (beyond selecting among equilibrium strategies)?

Exploitation algorithms

1.  Risk what you've won so far
2.  Risk what you've won so far in expectation (over nature's & own randomization), i.e., risk the gifts received
 - Assuming the opponent plays a nemesis in states where we don't know

...

- **Theorem.** A strategy for a 2-player 0-sum game is safe iff it never risks more than the gifts received according to #2
- Can be used to make any opponent model / exploitation algorithm safe
- No prior (non-eq) opponent exploitation algorithms are safe
- #2 experimentally better than more conservative safe exploitation algs
- Suffices to lower bound opponent's mistakes

STATE OF TOP POKER PROGRAMS

Rhode Island Hold'em

- Bots play optimally
[Gilpin & Sandholm EC-06, J. of the ACM 2007]

Heads-Up Limit Texas Hold'em

- Bots surpassed pros in 2008
[U. Alberta Poker Research Group]



- “Essentially solved” in 2015 [Bowling et al.]

Heads-Up No-Limit Texas Hold'em

Annual Computer Poker Competition



--> Claudico

- Statistical significance win against every bot
- Smallest margin in IRO: 19.76 ± 15.78
- Average in Bankroll: 342.49
(next highest: 308.92)

“BRAINS VS AI” EVENT

- Claudico against each of 4 of the top-10 pros in this game
- 4 * 20,000 hands over 2 weeks
- Strategy was precomputed, but we used endgame solving [Ganzfried & Sandholm AAMAS-15] in some sessions





Doug Polk @DougPolkPoker



BRAINS VS. ARTIFICIAL INTELLIGENCE

dougpolk7a 585 / 800

Player	Balance
DougPolk	-47091
Claudico_Polk	47091

- Hotkeys ▾
- Leave Match

#BrainsvsAI

Claudico_Polk
19750

Dealer

7♣ 10♣ 4♥
7 10 4

500

Actions: r250c/

DougPolk
19750

Min 0.5 0.75 Pot 2xPot All-in Wager Amount

100

Fold Check Bet to



TUOMAS SANDHOLM'S
ELECTRONIC
MARKETPLACES
LABORATORY

Carnegie Mellon University

RIVERS CASINO

GAMBLING PROBLEM? CALL 1-800-GAMBLER.



PITTSBURGH
SUPERCOMPUTING
CENTER

Humans' \$100,000 participation fee distributed based on performance



Overall performance

- Pros won by 91 mbb/hand
 - Not statistically significant (at 95% confidence)
 - Perspective:
 - Dong Kim won a challenge against Nick Frame by 139 mbb/hand
 - Doug Polk won a challenge against Ben Sulsky 247 mbb/hand
- 3 pros beat Claudico, one lost to it
- Pro team won 9 days, Claudico won 4

Observations about Claudico's play

- Strengths (beyond what pros typically do):
 - Small bets & huge all-ins
 - Perfect balance
 - Randomization: not “range-based”
 - “Limping” & “donk betting”
- Weaknesses:
 - Coarse handling of “card removal” in endgame solver
 - Because endgame solver only had 20 seconds
 - Action mapping approach
 - No opponent exploitation

Multiplayer poker

- Bots aren't very strong (at least not yet)
 - Exception: programs are very close to optimal in jam/fold games [Ganzfried & Sandholm AAMAS-08, IJCAI-09]

Conclusions

- Domain-independent techniques
- Abstraction
 - Automated lossless abstraction—exactly solves games with billions of nodes
 - Best practical lossy abstraction: potential-aware, imperfect recall, EMD
 - Lossy abstraction with bounds
 - For action and state abstraction
 - Also for modeling
 - Simultaneous abstraction and equilibrium finding
 - (Reverse mapping [Ganzfried & S. IJCAI-13])
 - (Endgame solving [Ganzfried & S. AAMAS-15])
- Equilibrium-finding
 - Can solve 2-person 0-sum games with 10^{14} information sets to small ϵ
 - $O(1/\epsilon^2) \rightarrow O(1/\epsilon) \rightarrow O(\log(1/\epsilon))$
 - New framework for fast gradient-based algorithms
 - Works with gradient sampling and can be run on imperfect-recall abstractions
 - Regret-based pruning for CFR
 - Using qualitative knowledge/guesswork
- Pseudoharmonic reverse mapping
- Opponent exploitation
 - Practical opponent exploitation that starts from equilibrium
 - Safe opponent exploitation

Current & future research

- Lossy abstraction with bounds
 - Scalable algorithms
 - With structure
 - With generated abstract states and actions
- Equilibrium-finding algorithms for 2-person 0-sum games
 - Even better gradient-based algorithms
 - Parallel implementations of our $O(\log(1/\epsilon))$ algorithm and understanding how #iterations depends on matrix condition number
 - Making interior-point methods usable in terms of memory
 - Additional improvements to CFR
- Endgame and “midgame” solving with guarantees
- Equilibrium-finding algorithms for >2 players
- Theory of thresholding, purification [Ganzfried, S. & Waugh AAMAS-12], and other strategy restrictions
- Other solution concepts: sequential equilibrium, coalitional deviations, ...
- Understanding exploration vs exploitation vs safety
- Application to other games (medicine, cybersecurity, etc.)

Thank you!

Students & collaborators:

- Noam Brown
- Christian Kroer
- Sam Ganzfried
- Andrew Gilpin
- Javier Peña
- Fatma Kılınç-Karzan
- Sam Hoda
- Troels Bjerre Sørensen
- Satinder Singh
- Kevin Waugh
- Kevin Su
- Benjamin Clayman

Sponsors:

- NSF
- Pittsburgh
Supercomputing Center
- San Diego
Supercomputing Center
- Microsoft
- IBM
- Intel