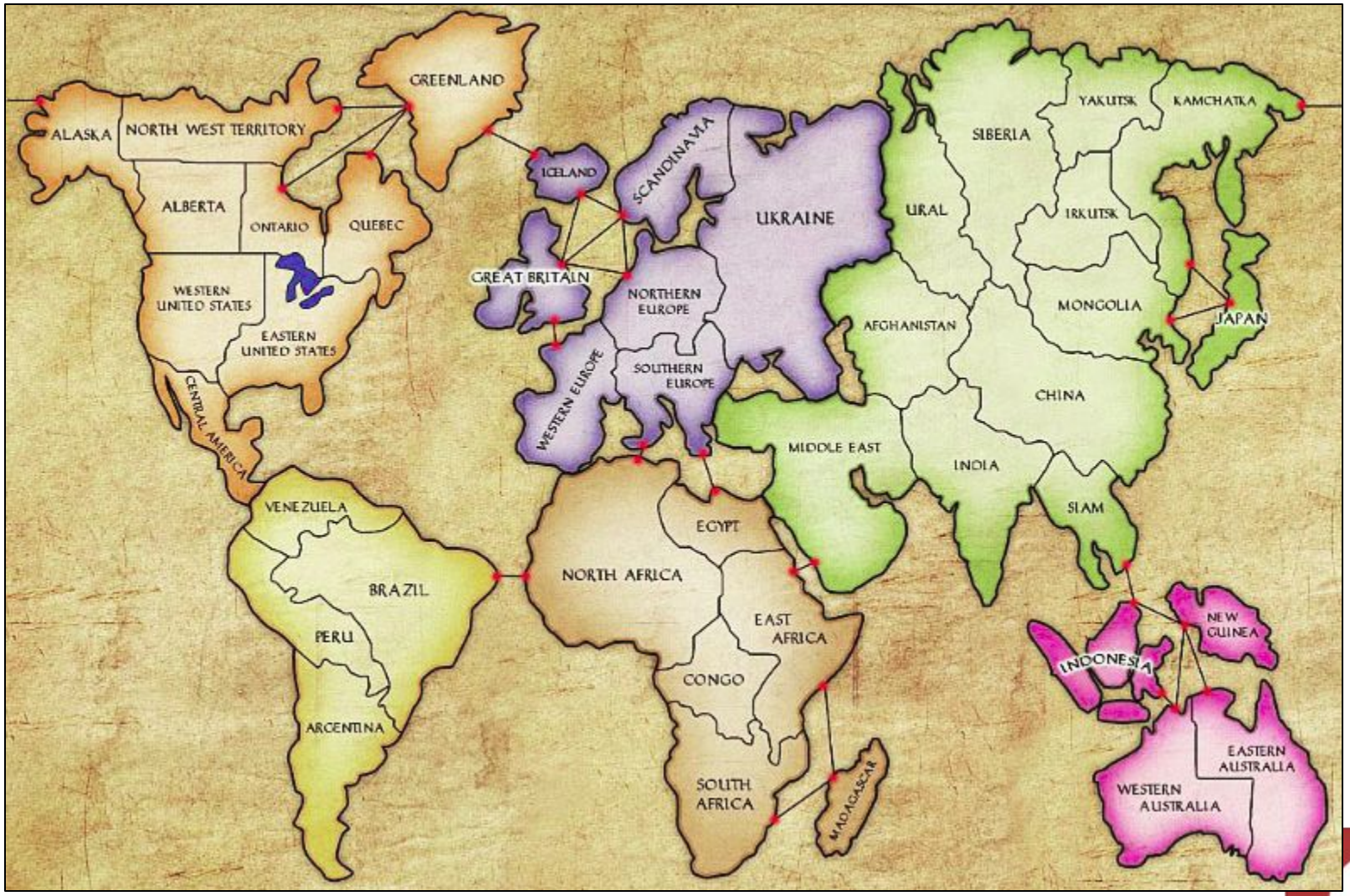
AI For World Domination

Oliver Krengel, Edward Terry, Henry Chen

April 23, 2018

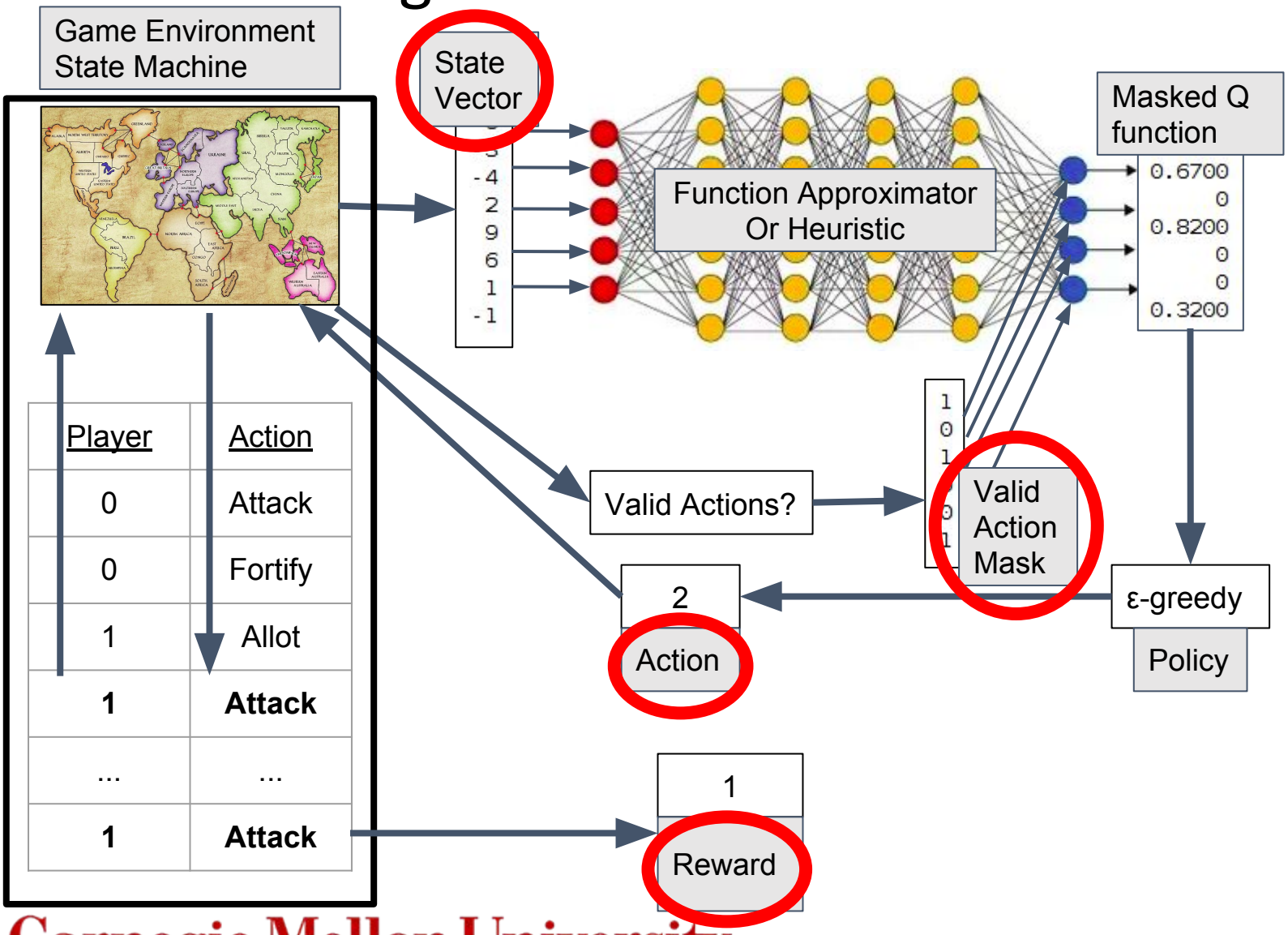
Risk: The board game



Challenges

- Environment Construction
  - Asynchronous turn order
  - Variable number of players
  - 4 Action types: Allot, **Attack**, Reinforce, Fortify
- State space
  - 42 territories
  - 7 players
  - Maximum 12 armies per territory
  - True state space: (12 * 7)^42 = 6.6e80
  - Simplified state space: (12 * 2)^42 = 9.3e57
  - 42 element vector
- Action space
  - 83 edges + 1 pass action = 84 element vector
  - Small subset are valid
- Highly stochastic dynamics
- State aggregation does not generalize well to valid actions
- But… reward shaping is straightforward!

Constructing The Markov Decision Process



Approach

Key decisions in construction:

- Flatten choices to single dimension - game theory
- **Game**: formulate object as state machine, naive to players
- **Environment**: translates game state into Markov Decision Process
- **Players**: objects that hold a policy for each action type
- **Policies**: common interface
- Built informed heuristics to deploy as adversaries, and for imitation learning

To train: Initialize with imitation learning, then play against itself

Algorithm for imitation learning:

1. Repeat for N games:
   a. Generate episode with heuristics: states (S), valid masks (V), actions (A)
   b. Add winner's (S,V,A) to dataset
2. Repeat for E epochs:
   a. Feed (S,V) into function approximator
   b. Compute cross-entropy loss
   c. Perform batch gradient descent

Hypothesis for Imitation Learning

Heuristics:

- *conservative* is typically better than *aggressive*, but not always
- Do not make any state-specific decisions
- Don't treat continents any differently

Hypothesis:

By using winner moves,

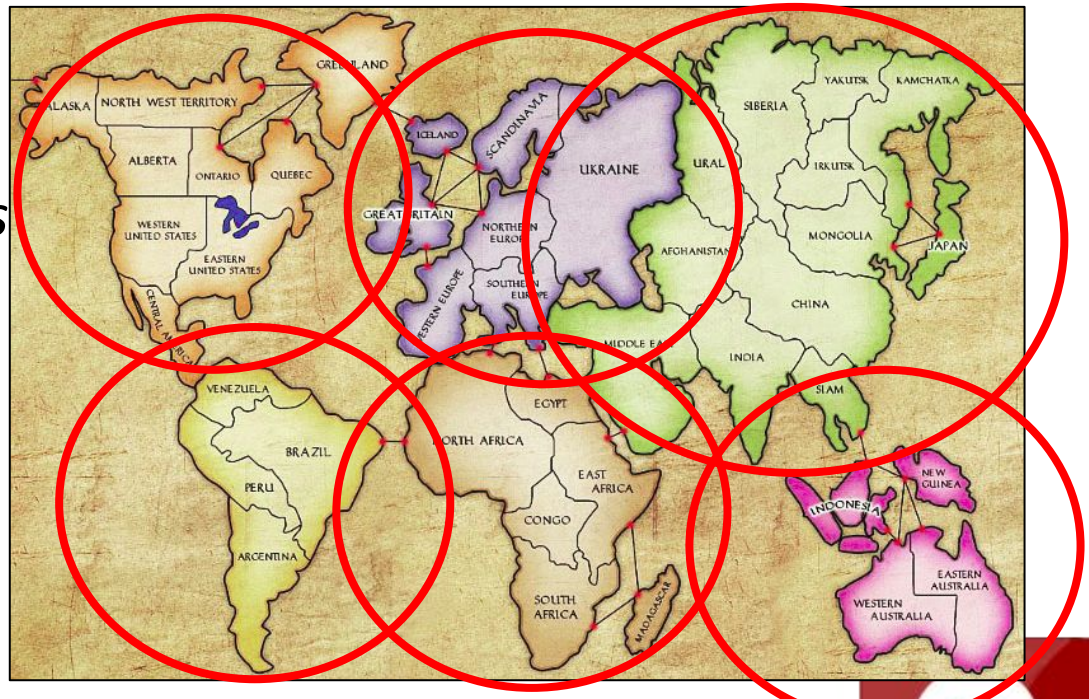
we will encode better policies

than the dominant heuristic

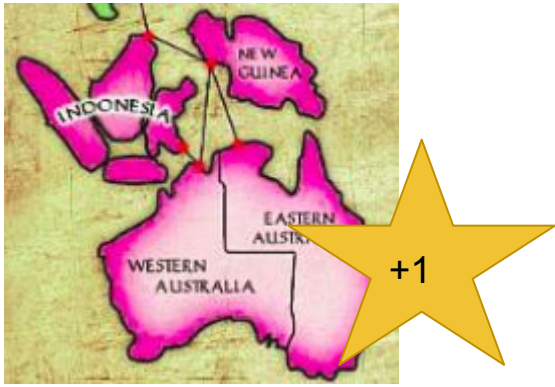
Reasoning:

Winners have made better

state-specific decisions



Results: Imitation Learning



Began with simpler environment: Australia!

Owner of Eastern Australia gets a bonus army

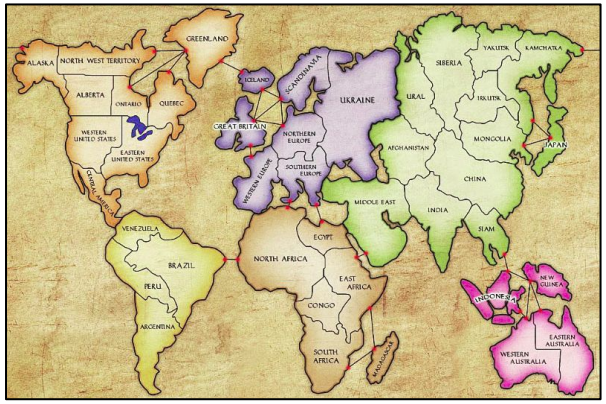
Hypothesis: best network will be #3

Network: {16:tanh:16:tanh:mask:sigmoid}

#	Hyperparameters					1000 games versus		
#	Players	Attack Heuristic	α	N	E	Conservative	Aggressive	Random
1	2	Conservative x 2	1e-4	1000	10000	506 wins	522 wins	733 wins
2	2	Conservative, Aggressive	1e-4	1000	10000	449 wins	451 wins	634 wins
3	2	Conservative, Aggressive	1e-4	1000	20000	518 wins	503 wins	668 wins
4	2	Aggressive x 2	1e-4	1000	1000	465 wins	488 wins	640 wins

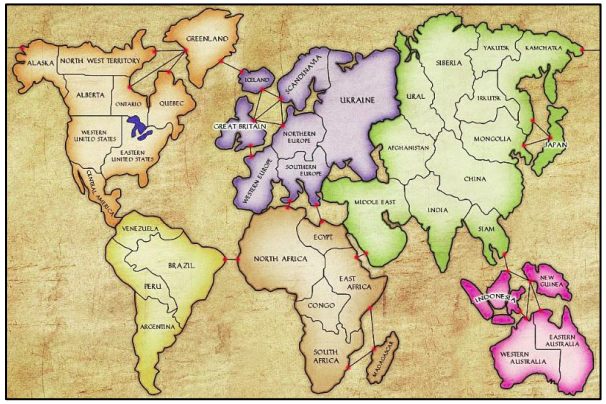
Results were inconclusive, but it took much longer to train the network attempting to learn from both policies

Full board challenges



- Generalizing human data: autoencoder?
- How to best explore state-space?
- Bad policies can still win
  - Random wins ~5% against 6 conservative opponents
  - Some untrained networks can beat random policies
  - Makes testing very difficult!
- Attack heuristics break down, “gloat”
- For game with 7 conservative players, states visited:
  - Median: 405, μ= 420, σ=102
  - Most of these states are end-game, irrelevant
  - Solution: trim number of states visited, using:
    - *Max_States* = $\mu - \sigma$ = 318
- Playing 100 games takes 250s on 4-core Intel i5 @2.5Ghz

Full board results



KEY: Attack policies

F	N	C	A	R
Untrained network	Trained network	Conservative heuristic	Aggressive heuristic	Random policy

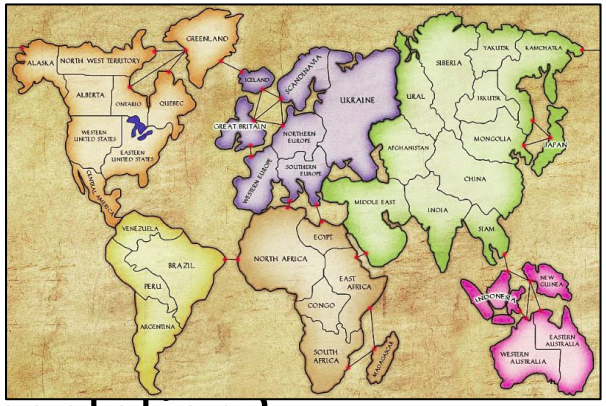
Hyperparameters

Architecture	α	N	E	Expert Policy	Max States
512:tanh:256:tanh:128	1e-4	100	2700	Winner of 7x conservative heuristics	318

Results: out of 100 games

	Wins by Player						
Matchup	**0**	**1**	**2**	**3**	**4**	**5**	**6**
RAAACCC	7	8	7	13	30	19	16
FAAACCC	2	7	5	6	29	30	21
NAAACCC	7	18	25	25	6	10	9
RC	6	94	-	-	-	-	-
NC	20	80	-	-	-	-	-

Lingering questions



- Train on/against human game data
- How to get state space sample
- More accurate n-player state space representation?
  - Separate maps for each player
  - Complex state-space representation
    - Angle corresponds to player
    - Magnitude corresponds to armies
    - Generalizes mid-game
- How to measure best policy?
  - No optimal game-theoretic policy exists
  - How can we manipulate other players??
- ……an utter lack of diplomacy……..