Part 2

A Perspective View and Survey of Meta-Learning

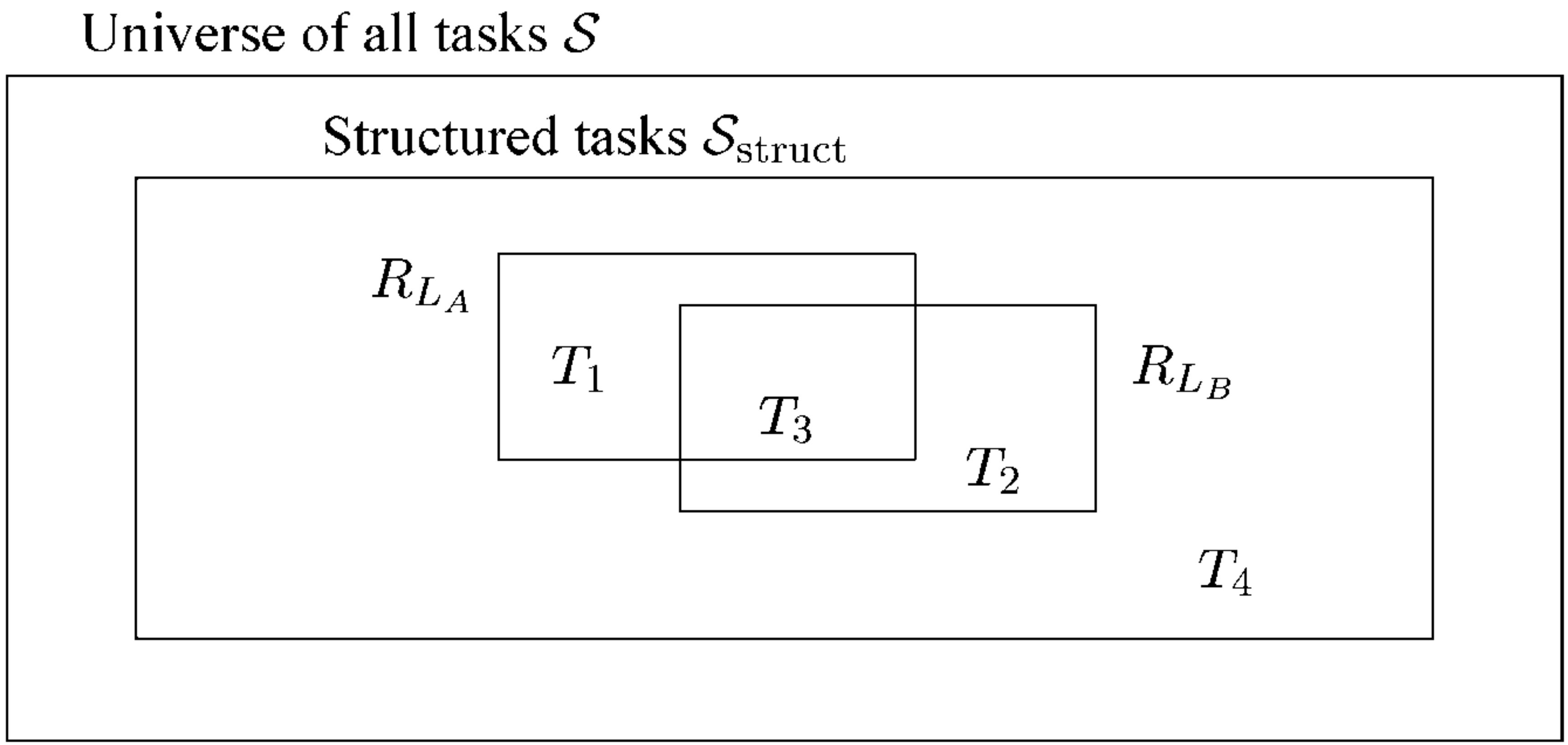
By Ricardo Vilalta and Youssef Drissi

Presented by Rasheed Ali R.

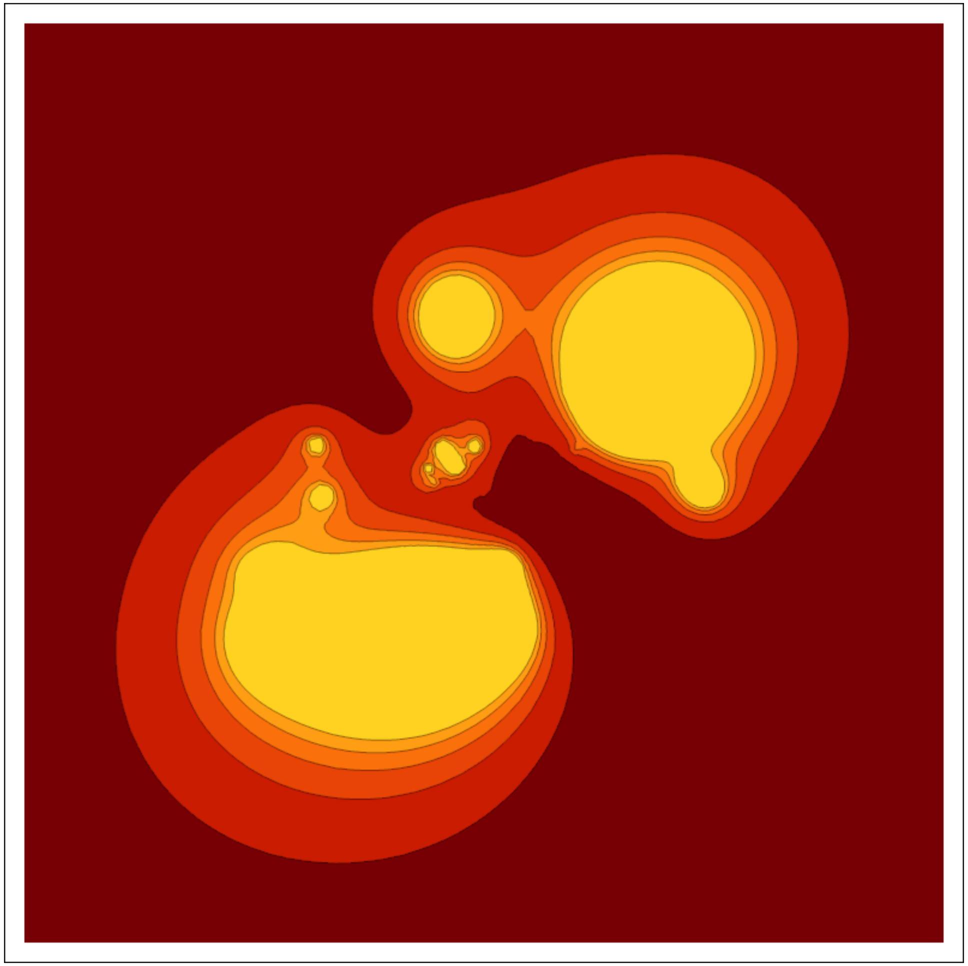
Discussion Type Approach

- Stacked Generalization is simple
- Interrupt with comments and questions
- No man behind

Base-Learning



Base-Learning



Solutions to Base-Learning Limits

- Direct application of Meta-Learning
  - Stacked Generalization
- Unbounded adaptive bias learners
  - Evolution Based Learners
  - Analogy Based Learners

Stacked Generalization by Example

- L_A, L_B, Γ
- $x = (x_1, x_2, x_3, x_4)$
- $x \in \{0,1,2,3\}$
- $c \in \{X, O\}$
- $T_{Train} = \{(x_i, c_i)\}_{i=1}^{m}$

Example Training Set

T_{train}	Samples	Class
(x_1, c_1)	$(1,2,1,0)$	X
(x_2, c_2)	$(0,1,2,1)$	O
(x_3, c_3)	$(3,0,0,2)$	X
(x_4, c_4)	$(1,2,3,1)$	X
(x_5, c_5)	$(2,1,0,2)$	O

$T_1 = \{(x_2, c_2$ … $c_5)\}$

$T_2 = \{(x_1, c_1$ … $c_5)\}$

$T_3 = \{(x_1, c_1$ … $c_5)\}$

$$T_4 = \{(x_1, c_1), (x_2, c_2), (x_3, c_3), (x_5, c_5)\}$$

$$T_5 = \{(x_1, c_1), (x_2, c_2), (x_3, c_3), (x_4, c_4)\}$$

Creating Meta-Training Set

$$h_{A1} = L_A(T_1),\ h_{B1} = L_B(T_1)$$

$$x_1' = (h_{A1}(x_1), h_{B1}(x_1))$$

T_{train}	Samples	Class
(x_1, c_1)	$(1,2,1,0)$	X
(x_2, c_2)	$(0,1,2,1)$	O
(x_3, c_3)	$(3,0,0,2)$	X
(x_4, c_4)	$(1,2,3,1)$	X
(x_5, c_5)	$(2,1,0,2)$	O

$$(x_i, c_i) \rightarrow (x_i', c_i)$$

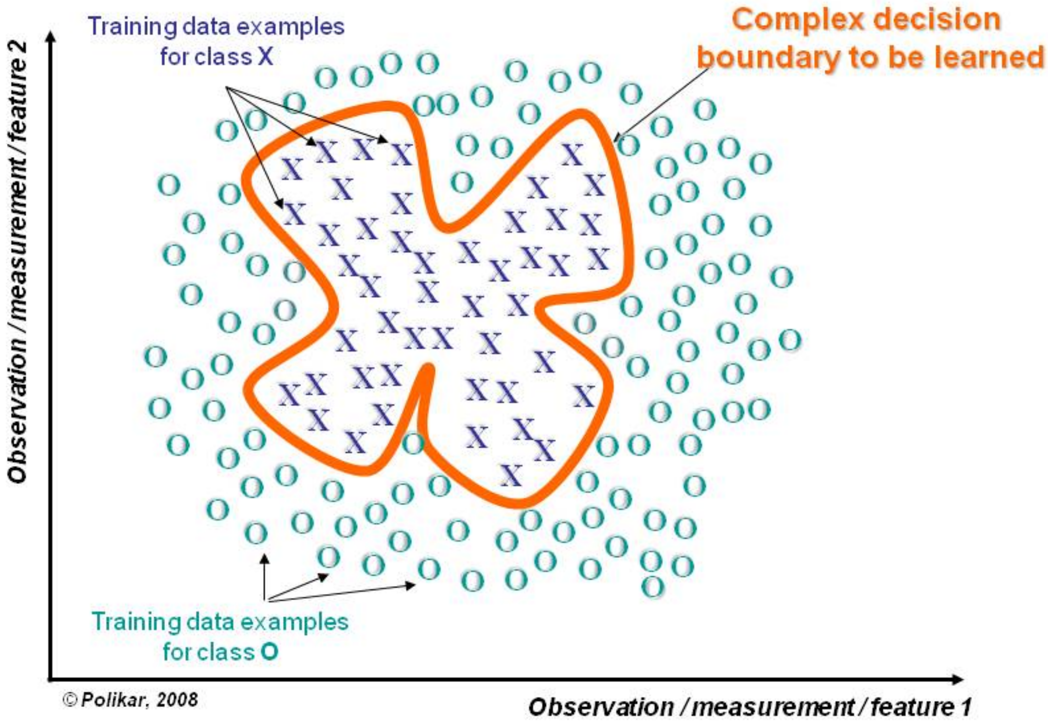
T_{train}'	Samples	Class
(x_1', c_1)	(X,X)	X
(x_2', c_2)	(O,O)	O
(x_3', c_3)	(X,O)	X
(x_4', c_4)	(O,X)	X
(x_5', c_5)	(O,O)	O

Meta-Learning

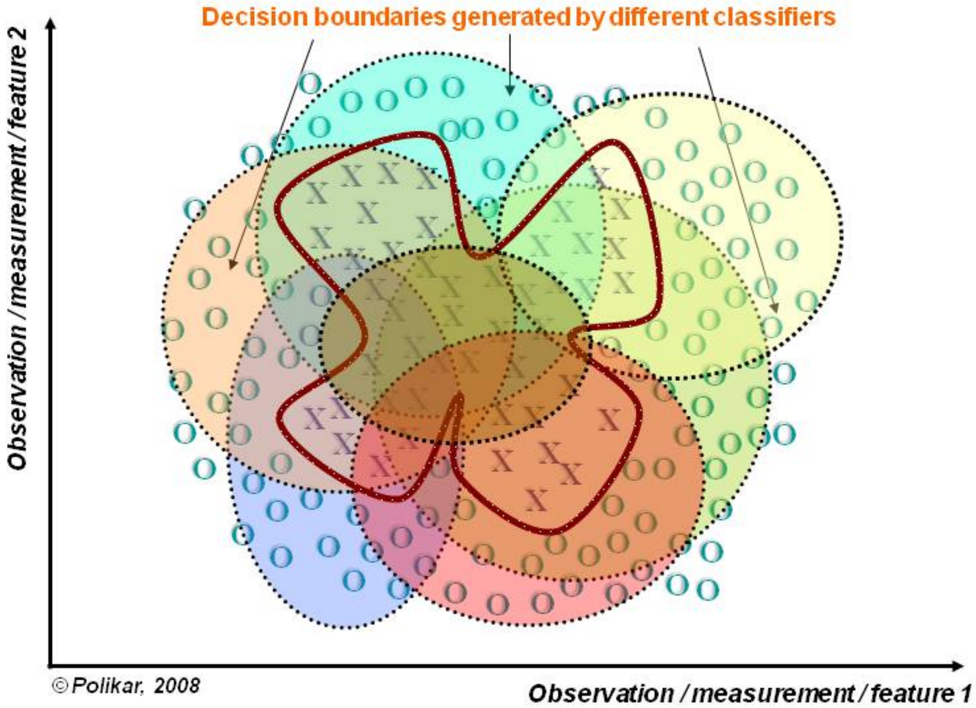
$$h=\Gamma(T_{train}^{\prime})$$

$$h(h_{A}(x),h_{B}(x))=c$$

Example


Training data examples
for class X
Complex decision
boundary to be learned
Observation / measurement / feature 2
Training data examples
for class O
© Polikar, 2008
Observation / measurement / feature 1

Example



Decision boundaries generated by different classifiers
Observation / measurement / feature 2
©Polikar, 2008
Observation / measurement / feature 1

Discussion

- How would you do it? What is wrong?
- What if all classifiers are correct
- What if all classifiers are wrong

T_{train}	Samples	Class
(x_1, c_1)	$(1,2,1,0)$	X
(x_2, c_2)	$(0,1,2,1)$	O
(x_3, c_3)	$(3,0,0,2)$	X
(x_4, c_4)	$(1,2,3,1)$	X

$$(x_i, c_i) \rightarrow (x_i', c_i)$$

T_{train}'	Samples	Class
(x_1', c_1)	(X,X)	X
(x_2', c_2)	(O,O)	O
(x_3', c_3)	(X,O)	X
(x_4', c_4)	(O,X)	X

Overview

- Learns from bias
- Predicts what is unpredicted by individual generalizers
- Generalizes winner takes all approaches
- Improves confidence
- Well suited for precision distributed and agent based systems

Meta-Learning Agents

- An Agent: $a{\downarrow}i$
  - Utility functions: $u{\downarrow}j\ (s\ )$
  - Decision function: $d(u\ )$
- Groups of Agents: $A=\{a{\downarrow}i\ \}$
  - diversity
  - $a{\downarrow}i\ (s\ )\rightarrow r$

Evolution Theory

1. Natural Selection:

An entity exists unless it parishes (Duh!)

$C{:}E \to E + \emptyset$

2. Reproduction:

Existing increases odds of replication

$R{:}E \to \{E,E\}$

3. Mutation:

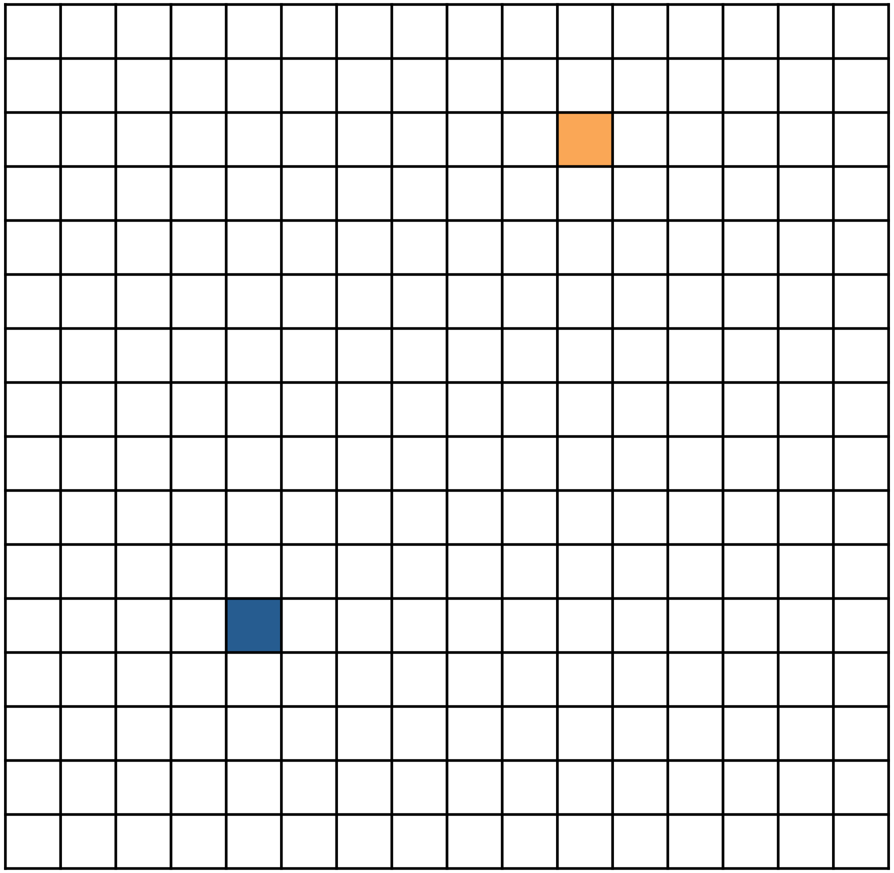
An entity may transform into another entity

$T(E) \to E\upharpoonleft$

Intelligence Agents & Evolution

- The pool of human thoughts:
  - C: A thought is forgotten unless it is minimally:
    - amusing
    - simple
    - useful
  - R: A thought gets replicated:
    - communication
  - T: A thought is modified
    - Accident: Misinterpretation/misrepresentation
    - Deliberate: Scientific or artistic process

Simple Evolutionary Learning



Agent Makeup

- General memory: 2040 Int64 cells
- Observation memory: 8 Int64 cells
  - Cell 0: remaining energy
  - Cell 1,2: x,y-coordinate of self
  - Cell 3,4: x,y-direction of self
  - Cell 5,6: x,y-coordinate of target

Agent Makeup

- Instruction Set:
  - Move: 100 energy
  - Turn: 5 energy
  - Copy: 1 energy
  - Think: 1 energy

Agent Makeup

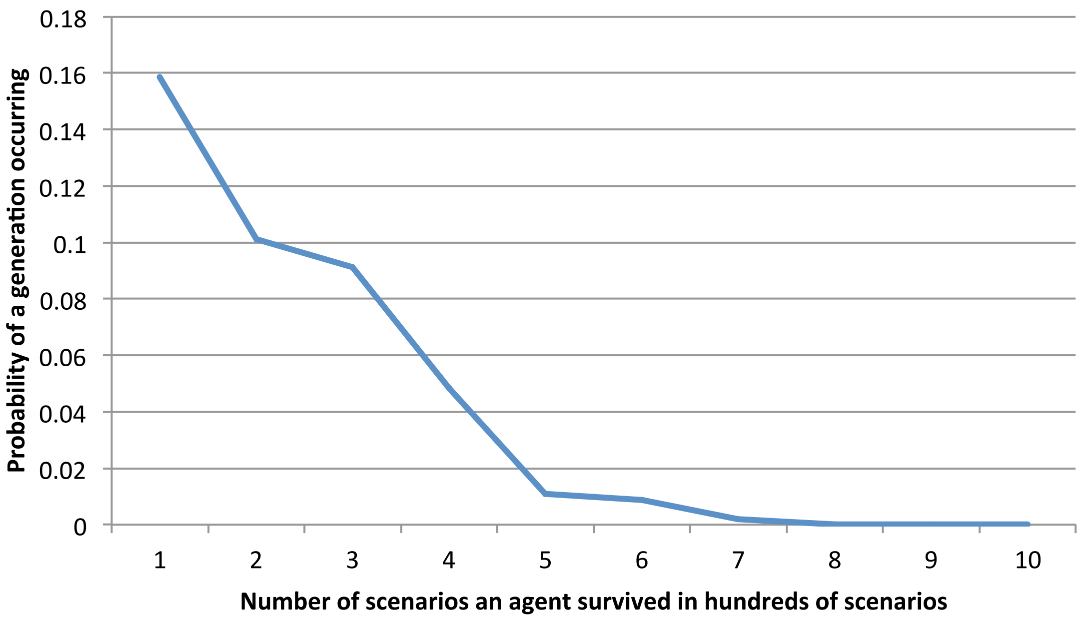
- Think instruction representation:
  - Instruction ID
  - Address of next instruction
  - Address of input start
  - Address of output start

Agent Thinking

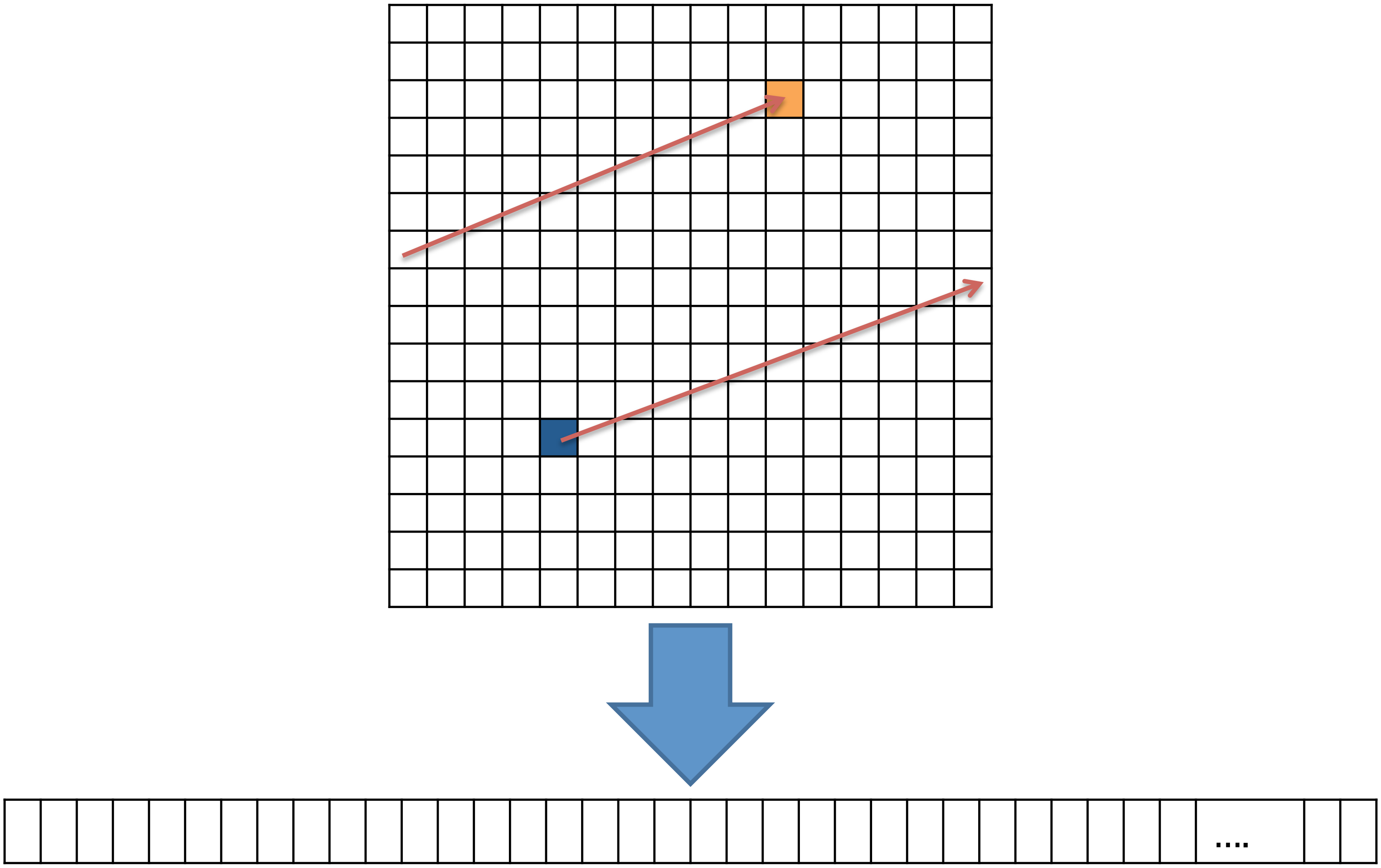
- Agents are Turing complete
- Thinking operation:
  - $Think(n,m) \to \pi_n(m)$ where $\pi_n \in S_k(\mathbb{N},\circ)$
- Thinking simulates any TM DFA δ
- Memory simulates TM tape/DFA states
- Copy action simulate all TM tape actions

Results

- 100,000 generations created
- Each generation evaluated 10,000 times
- Each evaluation determined agent survival
- Best agent survival probability: 0.0742



Best Static Strategy



Computations

- $Number\ of\ possible\ steps/total\ leniarized\ space = 20/256 = 0.078125$
- Any better results must be by thinking
- Out of the 10,000 generation some were thinking
- What are their traits

Improved Results

- Using a generation filtering tree
- On the order of $10 \uparrow 7$ million generations produced agents able to find 0.0803
- They must be doing better

Information Theory

Descriptive Complexity Theory

Analogy Learner Theory

Short Message Length Example