Research on a Learning Model Presuming Representationalism, Functionalism, and Neural Darwinism

Chang, Jen-Hao
Department of Electrical Engineering
National Taiwan University

Advisor: Yu, Tian-Li

Motivations

- Generalization is one of the most important abilities of humans.
- Instances + little prior knowledge ⇨ general rules
  - ✓ Humans
  - ? Learning model based on cognitive hypotheses

Cognitive Foundations

- Representationalism:

  the mind sees the world through representations

  ⇨ design knowledge representation

- Functionalism:

  the mind is composed of functions

  ⇨ design knowledge combination

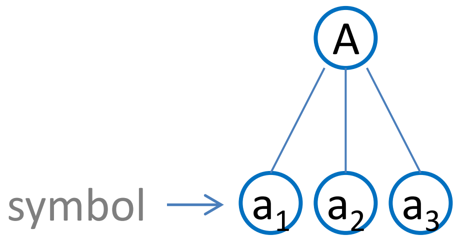
- Neural Darwinism:

  memory and learning are the result of evolution in brain

  ⇨ design learning algorithm

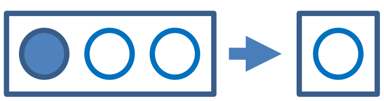
Knowledge Representation

- Based on representationalism
- Four types of representation
  - **Concept**: a group of things, with hierarchy
    ⇨ symbols, with hierarchy

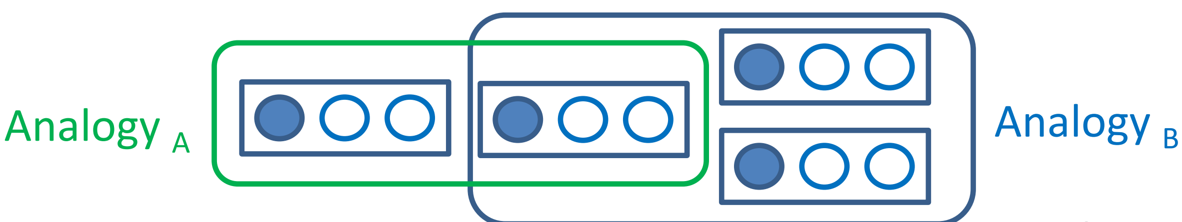


  - **Proposition**: a function of concepts
    ⇨ strings

  - **Rule**: function(inputs) → outputs
    ⇨ string → string



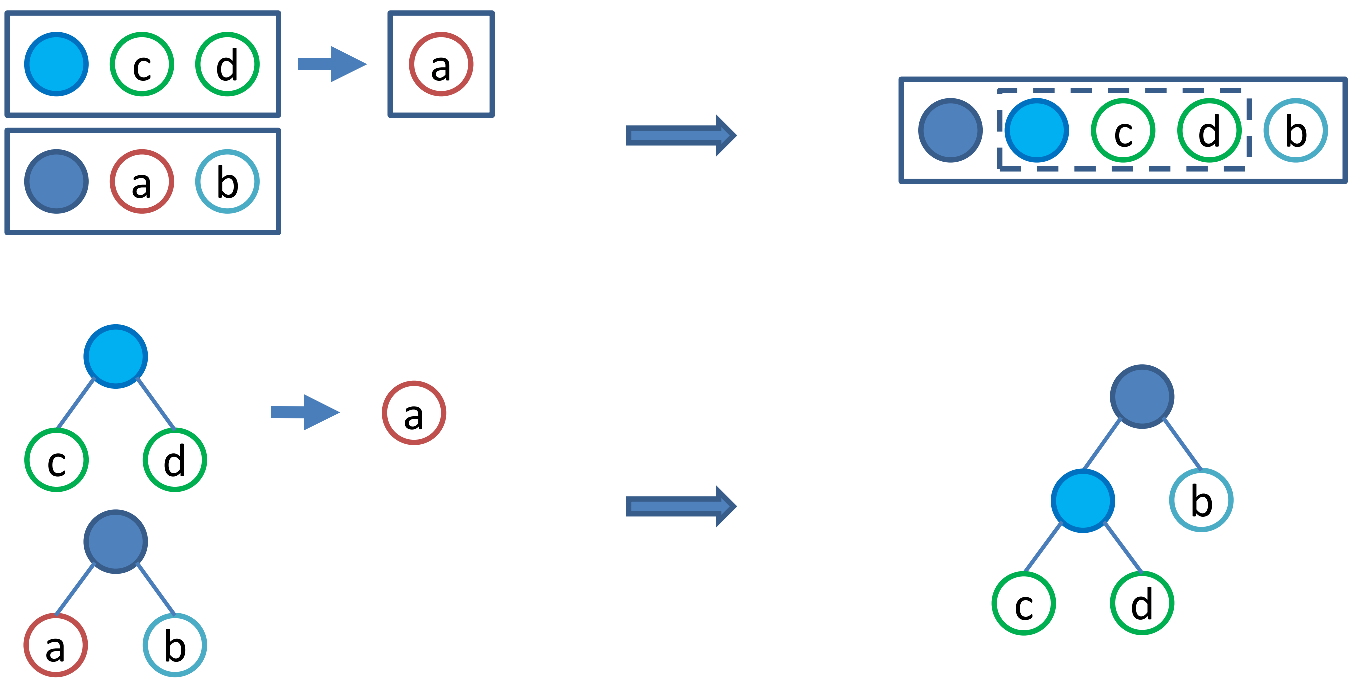
  - **Analogy**: the similarity between two situations
    ⇨ sets of strings



Knowledge Combination

- Based on functionalism
- The mind = a combination of functions

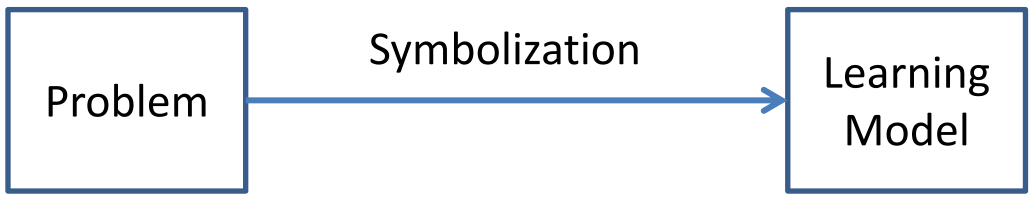
⇨ Propositions can combine with each other



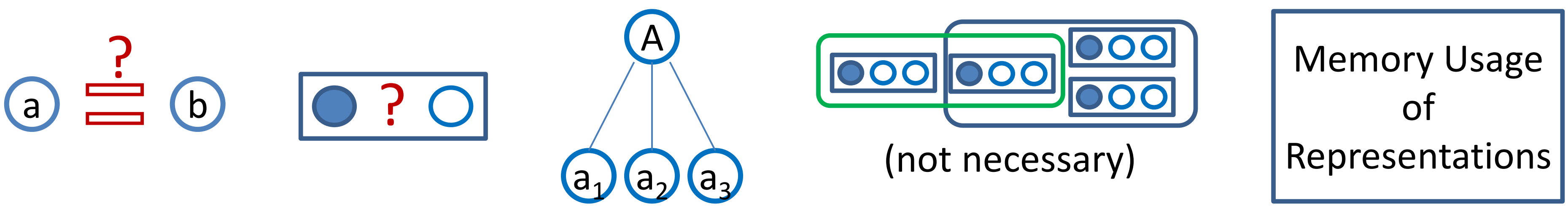
Learning Algorithm

- Based on neural Darwinism
  - selection upon variation of rule representations
  - Implemented with evolutionary algorithms

Prior Knowledge



Only bare-bone prior knowledge is given:



(1) (2) (3) (4) (5)

Assumptions:

1. The knowledge domain is deterministic
2. Input instances are consistent and their outputs are not functions

Fitness Function

- Adopt the minimum description length principle
- $m(a)$: the length of the encoded representation a
- $Fitness(r) = m(r) - \sum_{i \in C} m(i) + P \sum_{i \in I} m(i)$

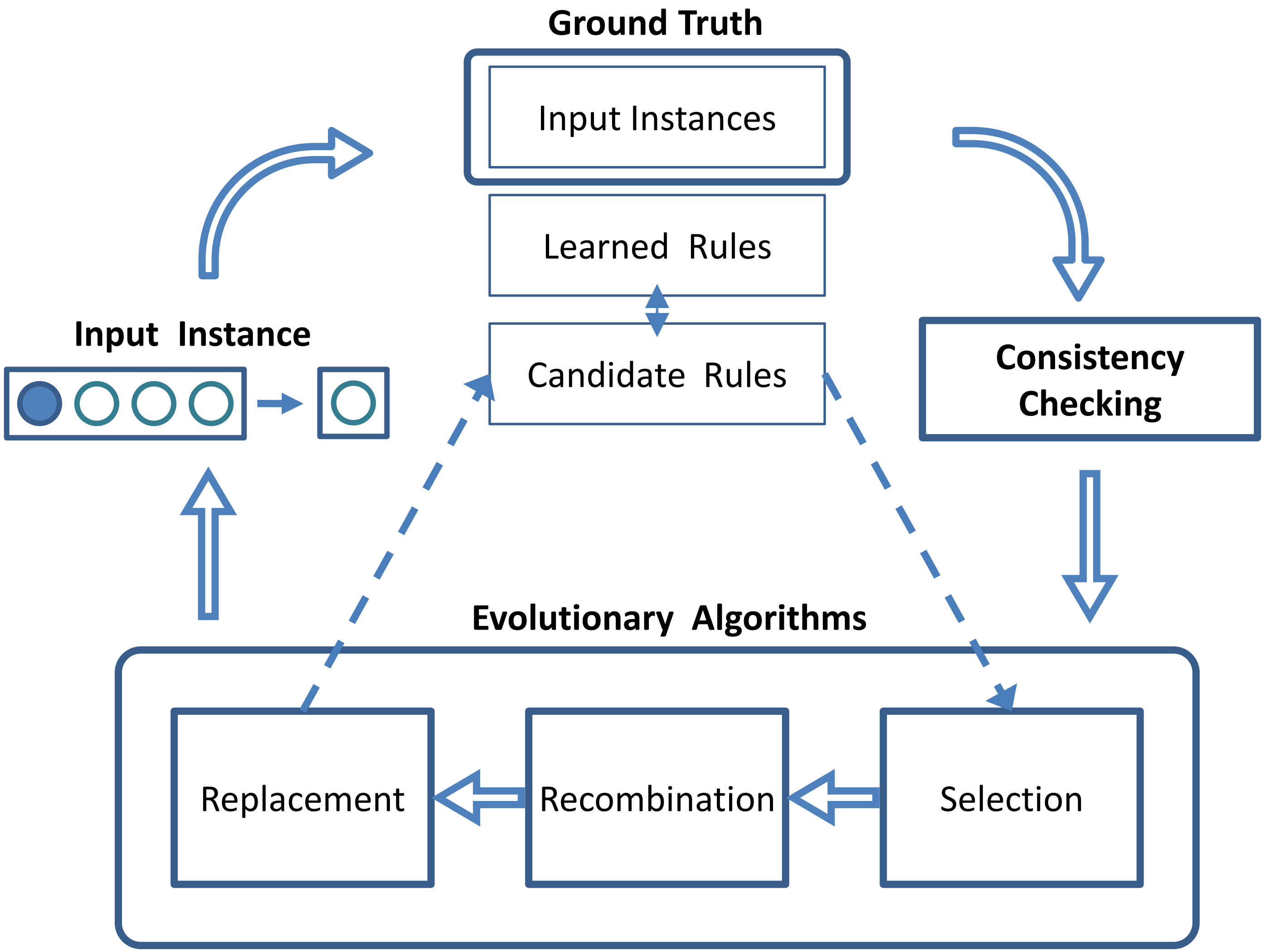
rule

Consistent ground instances

Inconsistent ground instances

- P : the punishment factor (use moderate values)

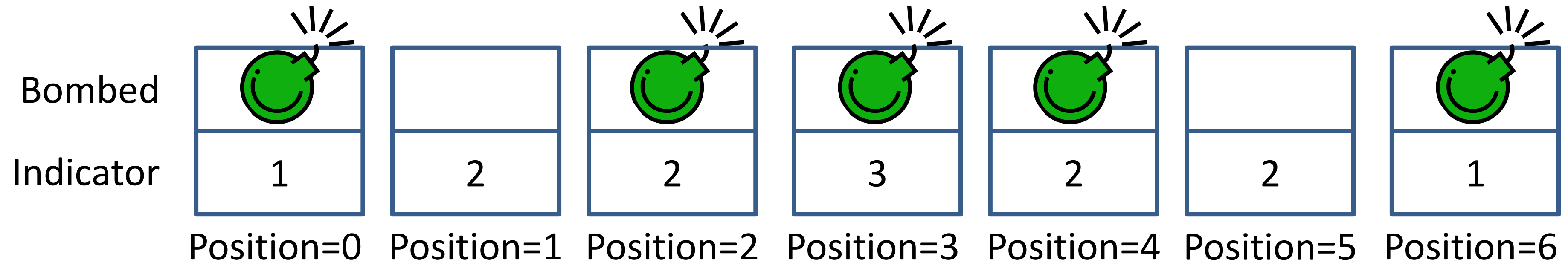
Overview of the Model



Analyses

- The model is Turing complete
  - The model ⇔ unrestricted grammar ⇔ Turing machine
- The learning process is *sound* but *incomplete*
  - The consistency of each rule is checked
  - The fitness function has bias
- The consistency checking is decidable
  - Use input instances to simplify rules

Experiment 1: 1-D minesweeper



Experiment 1: 1-D minesweeper

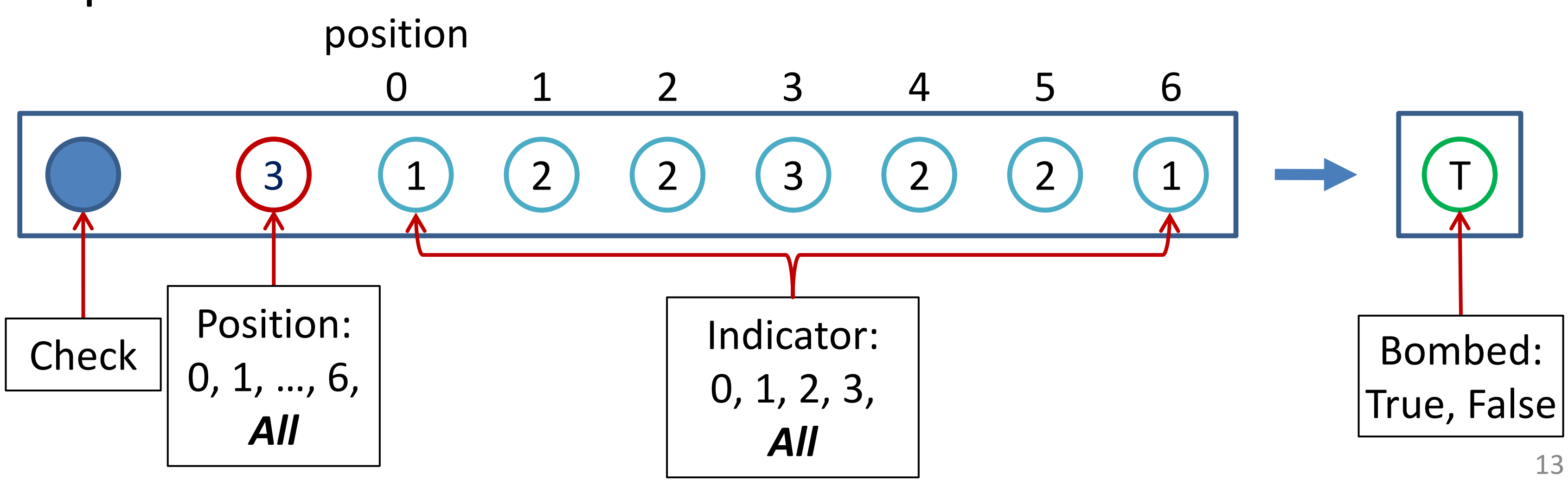
	Position=0	Position=1	Position=2	Position=3	Position=4	Position=5	Position=6
Bombed				?			
Indicator	1	2	2	3	2	2	1

- Given the indicators, is a cell at a position bombed ?

Experiment 1: 1-D minesweeper

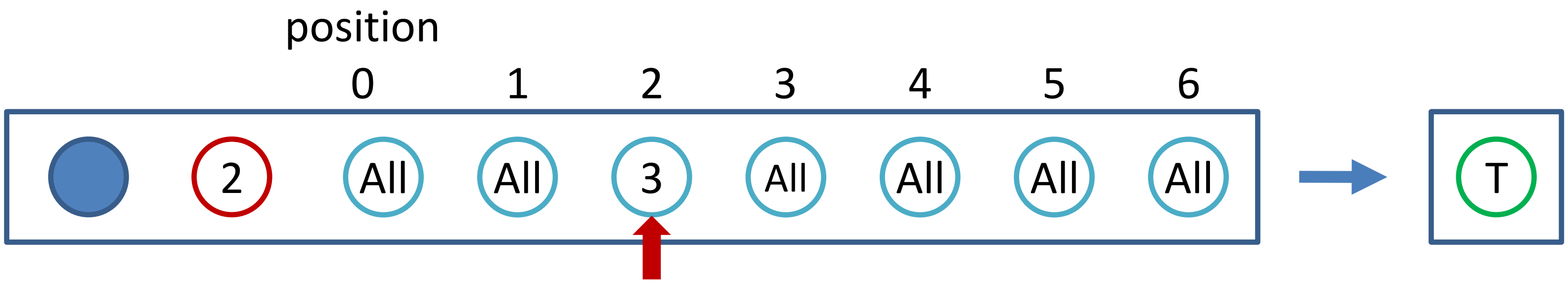
Bombed				**?**			
Indicator	1	2	2	3	2	2	1
	Position=0	Position=1	Position=2	Position=3	Position=4	Position=5	Position=6

- Given the indicators, is a cell at a position bombed ?
- Input Instance:



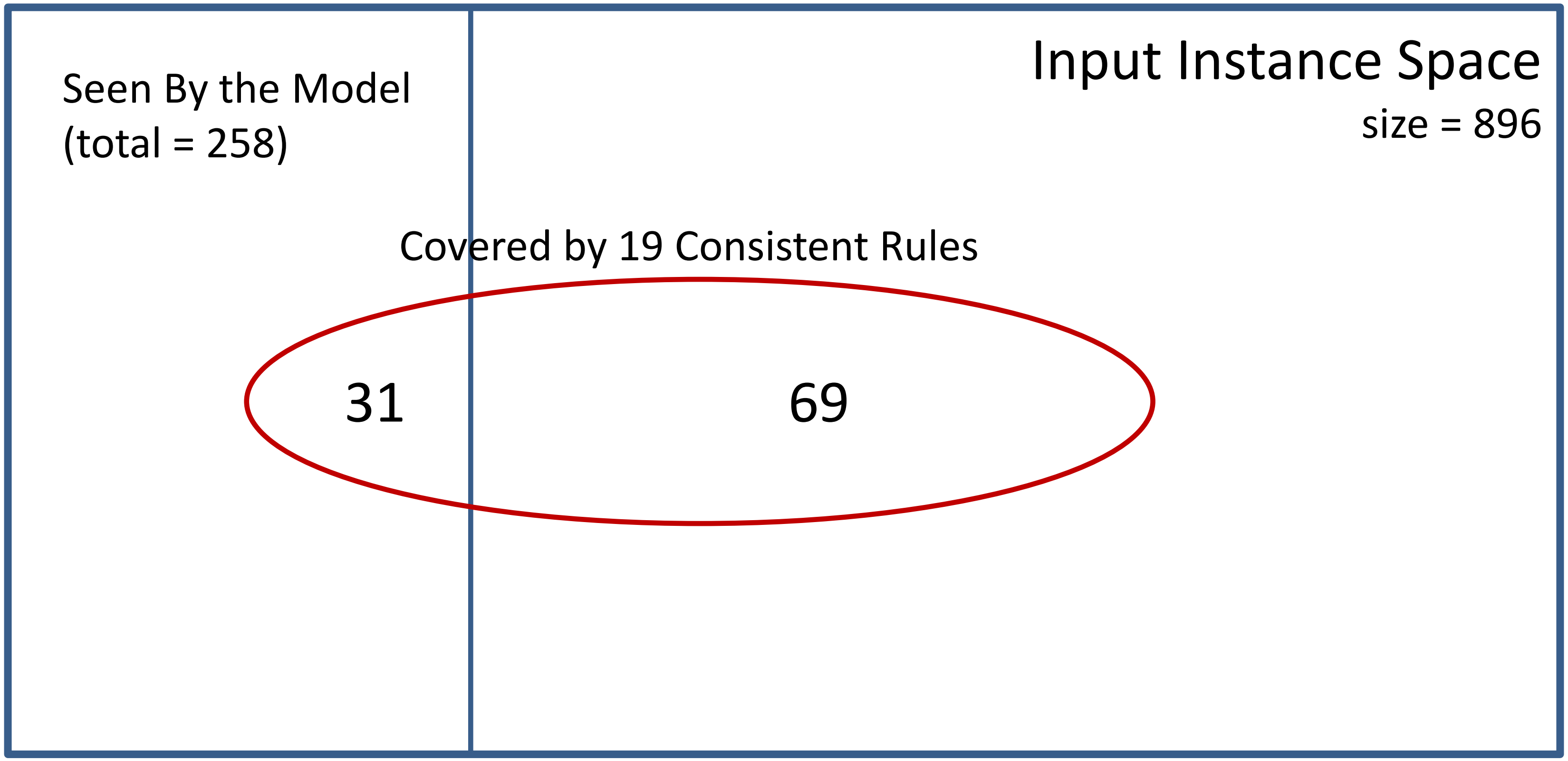
An Example

A consistent rule:



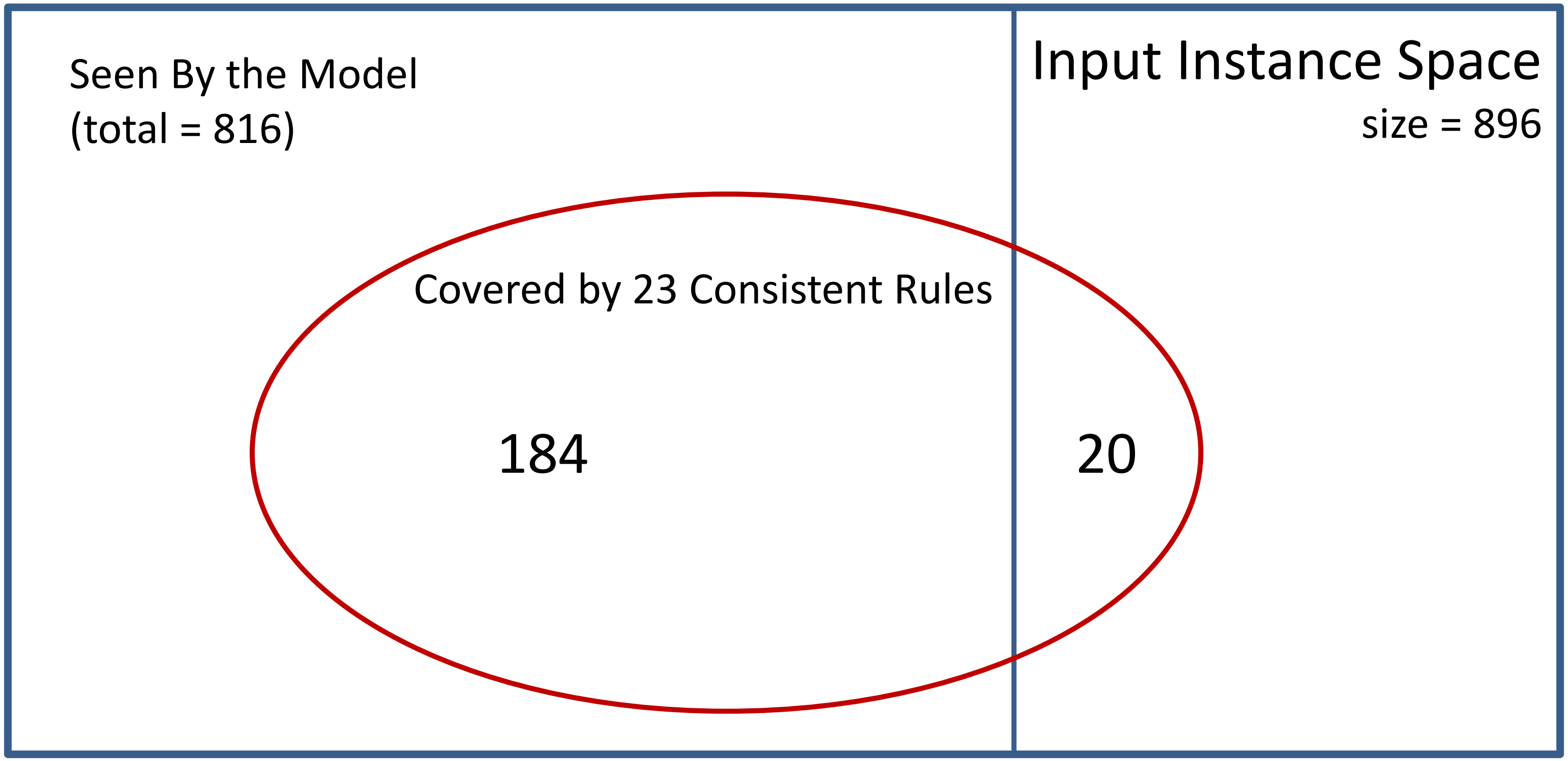
Results (1)

After 100 generations (population size = 25) :



Results (1)

After 2000 generations (population size = 25) :



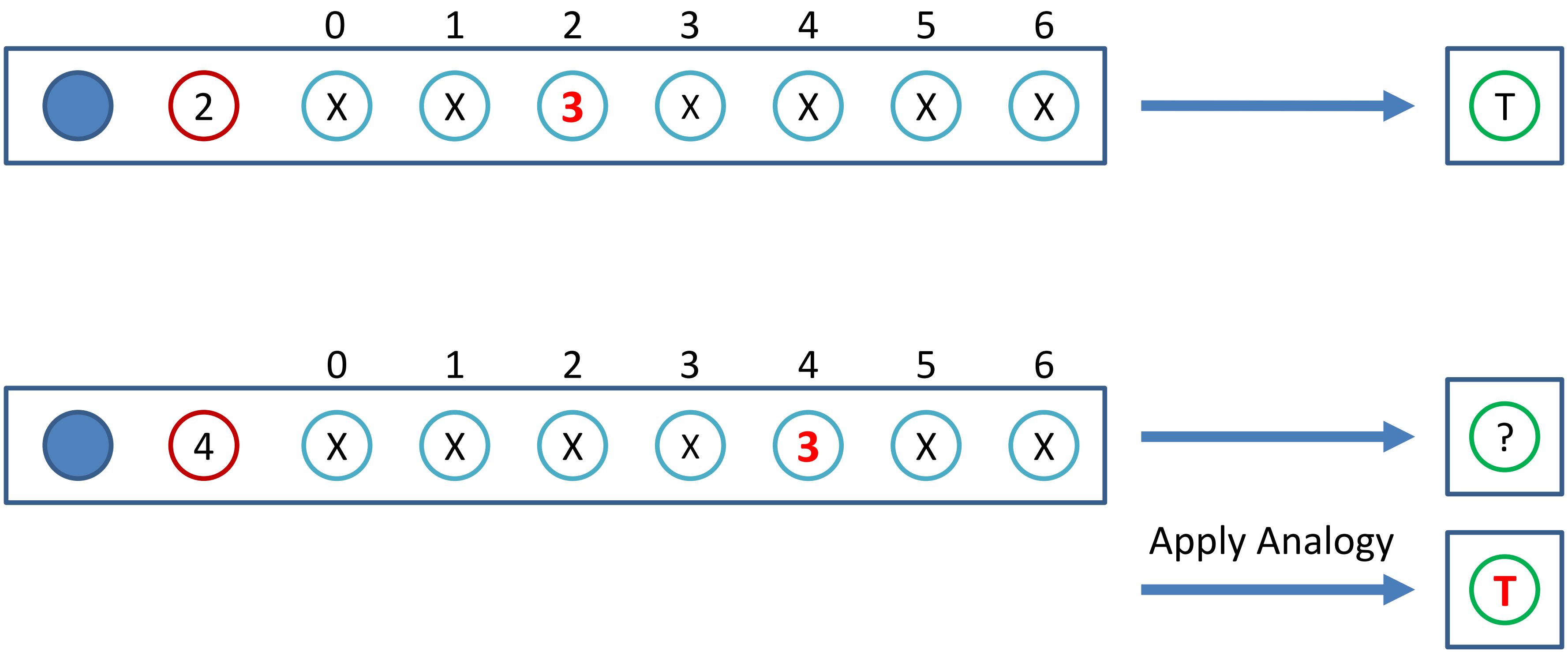
Results (2)

After 2000 generations (population size = 25) :

- 32550 different rules have been discovered by the model
  - 99.1% does not enter the population (filtered by the fitness function)
  - 14% are consistent (vs. random guess 2.6%)
- 281 different rules (0.9%) have entered the population
  - 56% are consistent

Experiment 1 + Analogy

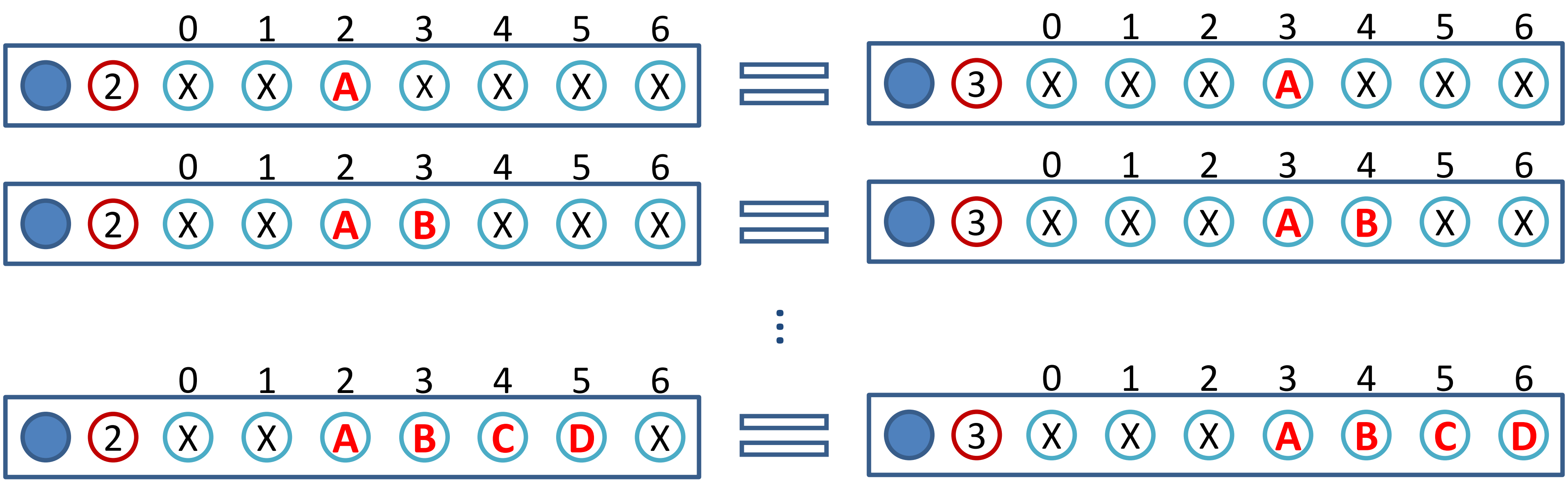
- Analogy allows the model to apply rules to similar situations



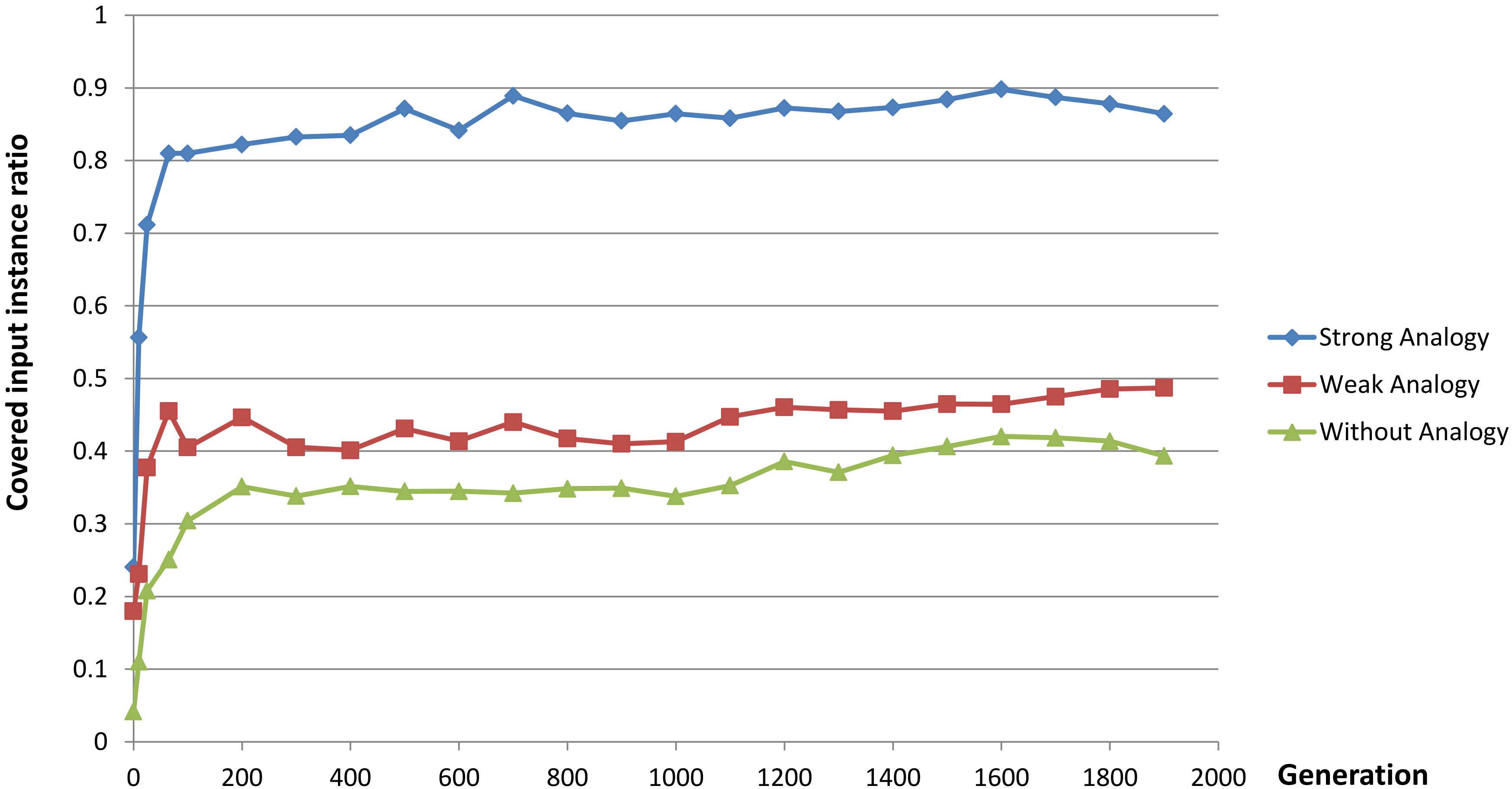
Two Settings of Analogies

- **Setting 1**: weak analogy (current position)

- **Setting 2**: strong analogy (current position + regions with size = 2, 3, 4)

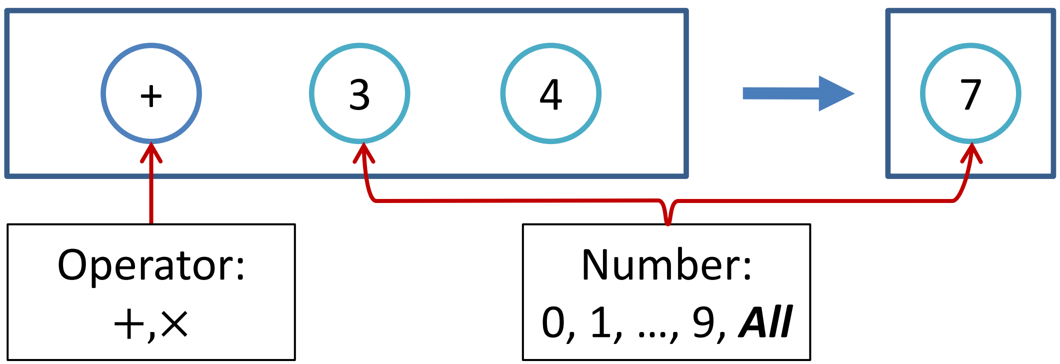


Results


Covered input instance ratio
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
200
400
600
800
1000
1200
1400
1600
1800
2000
Generation
Strong Analogy
Weak Analogy
Without Analogy

Experiment 3: mathematical ring

- Input instance:
  Integers with $+$, $\times$ (mod 10)



- Allow higher-level propositions to be generated

Results

After 2000 generations (population size = 25) :

- 5955 different rules have been discovered by the model
  - 93.1% does not enter the population (filtered by the fitness function)
  - 3.5% (208) are consistent
- 409 different rules (6.9%) have entered the population
  - 16% (65) are consistent
- 687 different instances are covered by 23 consistent rules in the population
  - 82% have higher level syntactical structures

Learning Results

Population Size	$0 \cdot N = N \cdot 0 = 0$	Identity of $\times$	Identity of +	$2 \times N = N \times 2 = N + N$	Commutativity of +	Associativity of $\times$	Associativity of +	Distributivity
25	1.00 (65)	1.00 (182)	1.00 (265)	0.65 (281)	0.60 (283)	0.00	0.00	0.00
100	1.00 (10)	1.00 (52)	1.00 (83)	1.00 (75)	1.00 (155)	0.10 (1223)	0.05 (1072)	0.00
200	1.00 (4)	1.00 (14)	1.00 (34)	1.00 (33)	1.00 (88)	0.24 (1228)	0.24 (1342)	0.00
500	1.00 (0)	1.00 (6)	1.00 (4)	1.00 (22)	1.00 (28)	0.50 (921)	0.33 (700)	0.00

Summary

- A learning model based on three cognitive hypotheses
- Only bare-bone prior knowledge is given
- The knowledge base speeds up the consistency checking
- The model is analyzed:
  1. Turing complete
  2. sound, incomplete
  3. consistency checking is decidable
- The fitness function:
  1. based on the minimum description length principle
  2. filter effect
- Three experiments:
  1. generalization
  2. analogy can enhance the generalization ability
  3. rules with unseen syntactical structures can be learned

Conclusion

- A learning model based only on the three cognitive hypotheses can do inductive learning.
- Even provided only with bare-bone prior knowledge, this learning model still can learn general rules from input instances.
- This learning model can utilize analogical information to enhance its learning ability.
- The model can learn general rules which have unseen syntactical structures.