Invariant Generation.

Ashish Kumar

January 25, 2022

Summary

1. Heuristics list for Cost Function.
2. An Example of a Cost Function adopting continuous learning from Z3.
3. Heuristics for Mutating Invariant.

Heuristics for Cost Function.

Used heuristics by earlier work:

- "Smaller values" are better features and better thresholds.
- Choose features which maximize distance from TS and classifier.[???]
- Given features, choose threshold which give more skewed partitions.

New Ideas:

- $I \wedge \neg B$ must be "close" to Q.
- Area of set not satisfying constraint - not sure how to compute this.
- Iterative converging methods.
  - Are there iterative solution finders to CHC equations - do they define an error function?
  - Get an ϵ-approximation to P, B, Q, and call a solution for this an ϵ- invariant.
- Learning from Z3[1].
  - Extract a continuous function from our learner (Z3) rather than a binary function.
  - Get k counterexamples. Define a continuous function over these k examples (area/ centroid/ distance).

[1] Need to figure out a way to mutate invariants

Learning from Z3 using s counterexamples: One Example.

Let the equations be (universally quantified over s,t):

$$P(s) \to I(s) \tag{1}$$
$$I(s) \wedge B(s) \wedge T(s,t) \to I(t) \tag{2}$$
$$I(s) \wedge \neg B(s) \to Q(s) \tag{3}$$

Let $J(I)$ be the cost function for a guess invariant. Then:

$$J(I) = K_1 J_1(I) + K_2 J_2(I) + K_3 J_3(I)$$

where $J_i(I)$ is the cost for violating equation (i). Get atmost s (try $s = 10$) counterexamples for each case.[2] Let z_i denote the ith counterexample for case 1 and case 3 and (z_i^1, z_i^2) for case 2.

$$J_1(I) = \max_{1 \le i \le s} d(z_i, I)$$
$$J_2(I) = \text{Penalty}(TS, I) \qquad (?)$$
$$J_3(I) = \max_{1 \le i \le s} d(z_i, Q)$$

For $J_2(I)$ the 'variance' of the ICE examples seems like a good fit[3].

[2] There will either be 0,1 or ∞ counterexamples for linear boundaries for (1) and (3).

[3] Computing $J_1(I)$ and $J_3(I)$ will require a method for non linear boundaries.

Mutating Invariants: Heuristics.

- In each iteration, randomly change some parameter of the invariant. If it decreases the cost, go to this invariant, else do not.
- If we fix the number of hyperplanes (to say k), then invariants close to $I(\theta_{11}, \theta_{12}..\theta_{1n}, ..\theta_{k1}, \theta_{k2}..\theta_{kn}, c_1, ..c_k)$ is $I(\theta_{11} \pm \epsilon, \theta_{12} \pm \epsilon..\theta_{1n} \pm \epsilon, ..\theta_{k1} \pm \epsilon, \theta_{k2} \pm \epsilon..\theta_{kn} \pm \epsilon, c_1 \pm \epsilon, ..c_k \pm \epsilon)$. But how to do this when k is variable? Other thought - should we do translations first, then rotations?