The Practical Assessment of Test Sets with Inductive Inference Techniques

Neil Walkinshaw

Department of Computer Science
University of Leicester

September 4, 2010

Background

Test Adequacy

- Assessing the ability of a test set to identify faults
  - Successful execution of an adequate test set should imply that there are no faults in a tested program
- How do you know if a test set is adequate?
- Numerous *adequacy criteria* have been developed
  - Statement / branch / path / data-flow, ...

Background

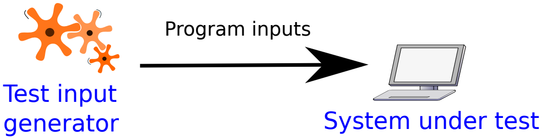
Test Adequacy

- Assessing the ability of a test set to identify faults
  - Successful execution of an adequate test set should imply that there are no faults in a tested program
- How do you know if a test set is adequate?
- Numerous *adequacy criteria* have been developed
  - Statement / branch / path / data-flow, …

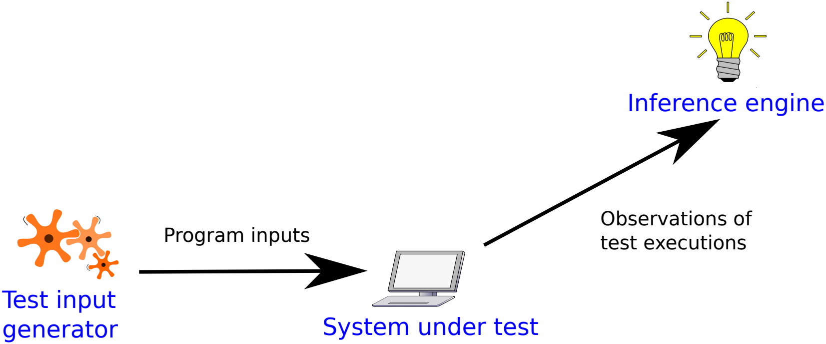
Problem

- Criteria based on syntax are often a poor approximation for actual adequacy

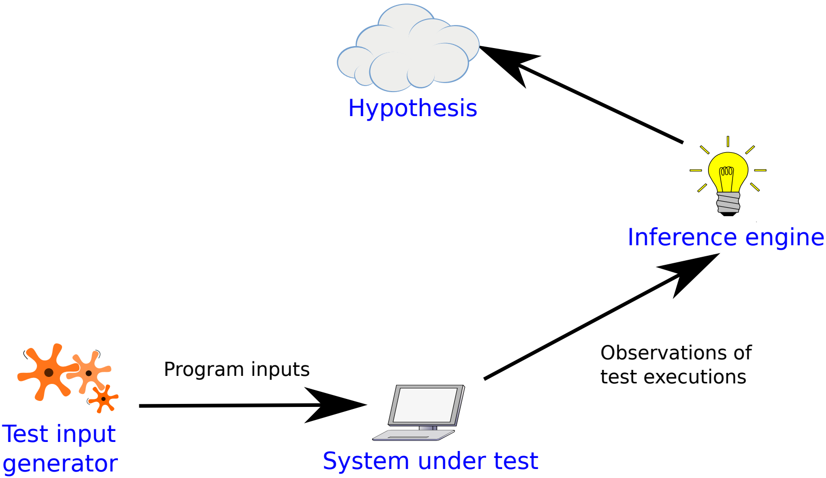
Using Inference to Assess Test Set Adequacy



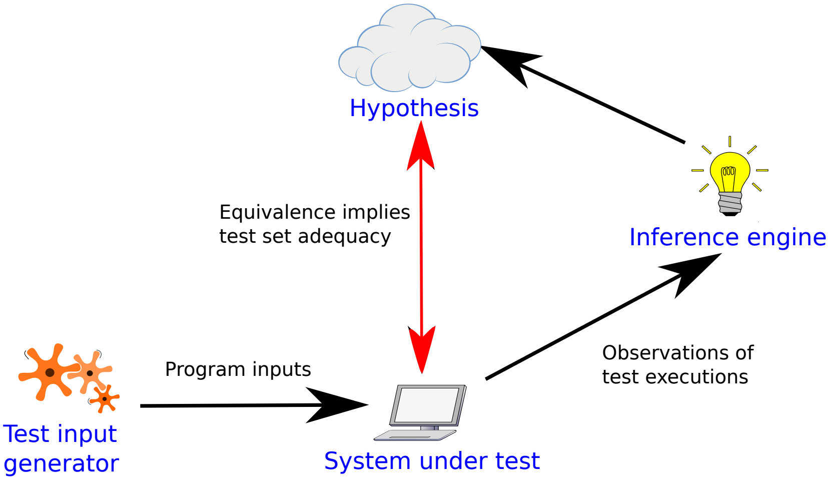
Using Inference to Assess Test Set Adequacy



Using Inference to Assess Test Set Adequacy

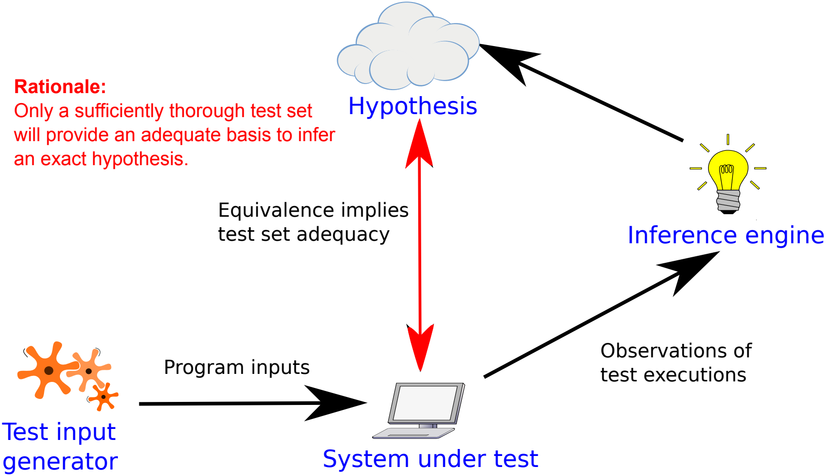


Using Inference to Assess Test Set Adequacy



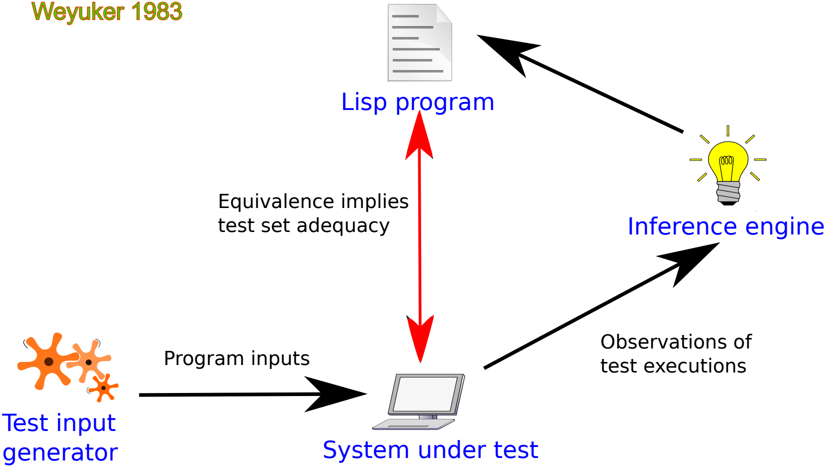
Using Inference to Assess Test Set Adequacy

Rationale:
Only a sufficiently thorough test set will provide an adequate basis to infer an exact hypothesis.



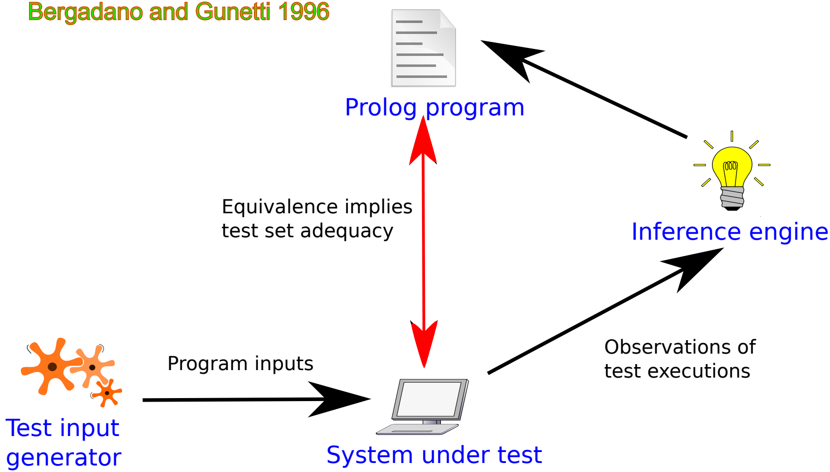
Using Inference to Assess Test Set Adequacy

Weyuker 1983



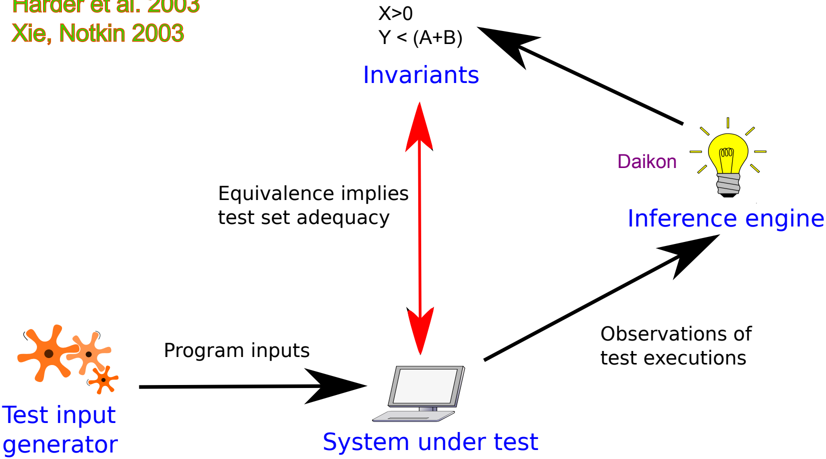
Using Inference to Assess Test Set Adequacy

Bergadano and Gunetti 1996



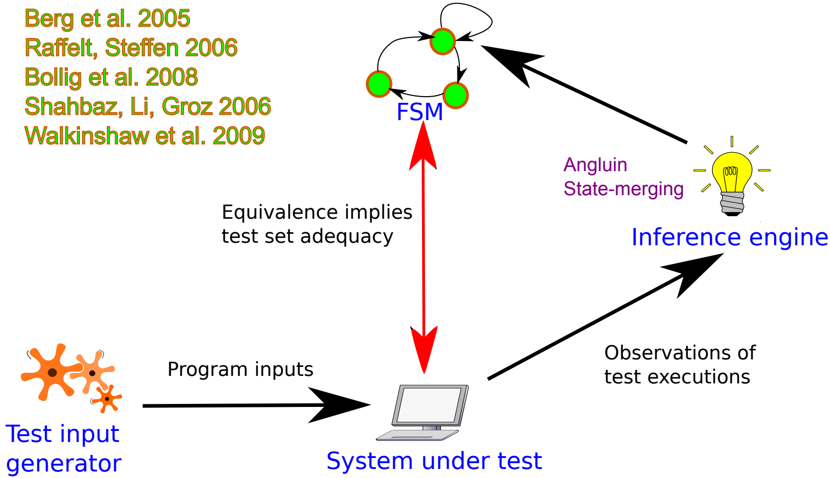
Using Inference to Assess Test Set Adequacy

Harder et al. 2003
Xie, Notkin 2003

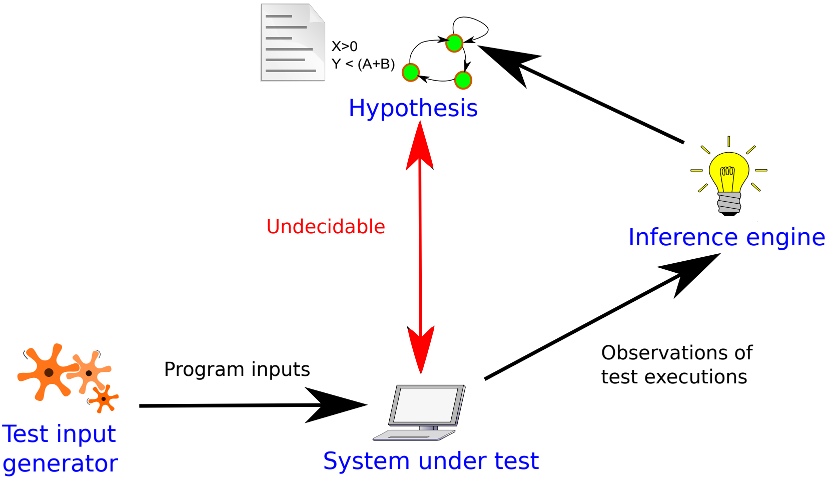


Using Inference to Assess Test Set Adequacy

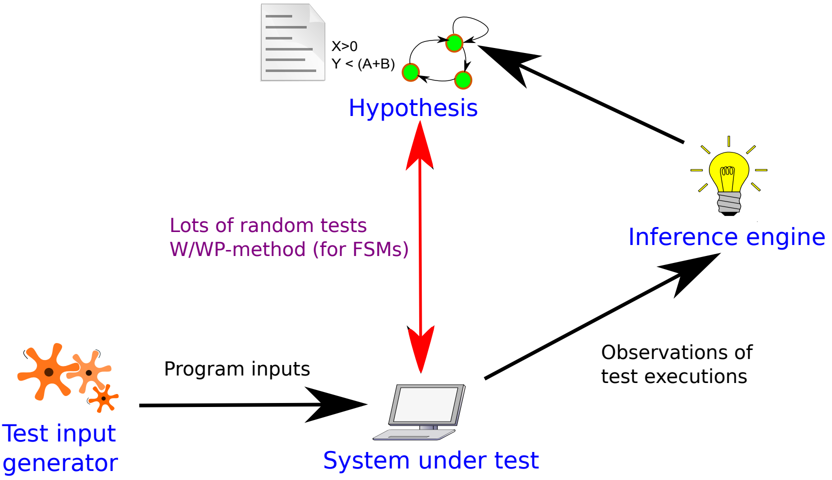
Berg et al. 2005
Raffelt, Steffen 2006
Bollig et al. 2008
Shahbaz, Li, Groz 2006
Walkinshaw et al. 2009



USING INFERENCE TO ASSESS TEST SET ADEQUACY



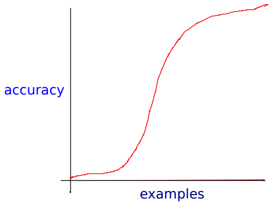
Using Inference to Assess Test Set Adequacy



Problem

Based on exact results - no flexibility

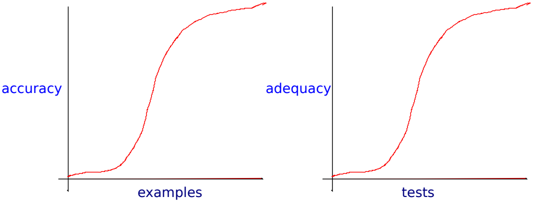
- The inferred model is either equivalent to the subject system or not.
    - The corresponding test set is either adequate or not.
- In reality, there is bound to be a certain degree of error.
    - A test set may result in a model that is 99% correct, with only small, trivial errors



Problem

Based on exact results - no flexibility

- The inferred model is either equivalent to the subject system or not.
  - The corresponding test set is either adequate or not.
- In reality, there is bound to be a certain degree of error.
  - A test set may result in a model that is 99% correct, with only small, trivial errors



The Probably Approximately Correct (PAC) framework

Setting

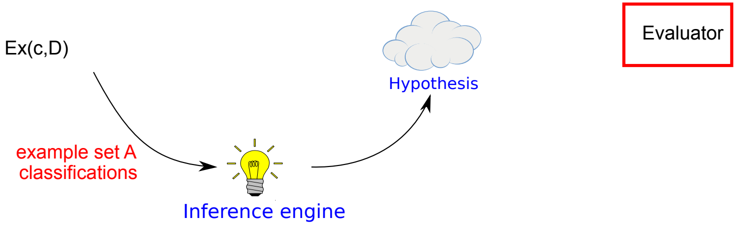
- There exists an *instance space* X
- The learning target is a *concept* $c \subset X$
  - For any element $x \in X$, $c(x) = 1$ or 0
- There is a *selection procedure* $EX(c, D)$ that randomly selects elements in X
  - The probability of them belonging to c is determined by some static distribution D (not necessarily known)
- Given a labelled set of examples selected by EX, it is the goal of the learning procedure to infer c

The Probably Approximately Correct (PAC) framework

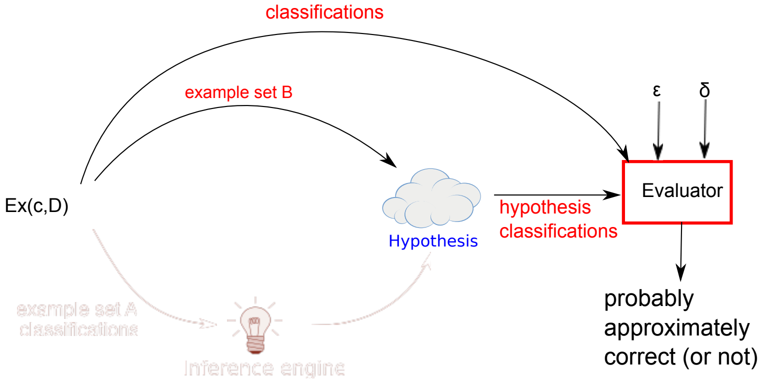
Assessing a Learner

- Two problems
  1. Can only guarantee accurate result if supplied with every possible instance in X.
  2. Given that samples are a random subset, there is the chance that EX will supply a misleading sample.
- To address these issues, the success of a learner is characterised as follows:
  - δ - probability that the hypothesis will meet the success conditions
  - ε - allowable degree of error

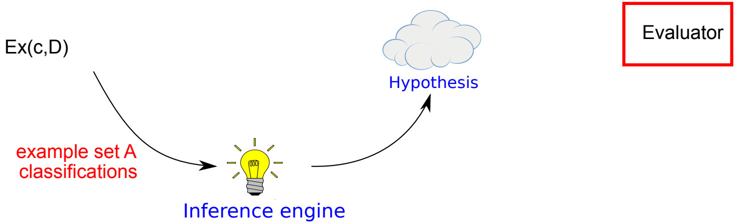
THE PROBABLY APPROXIMATELY CORRECT (PAC) FRAMEWORK



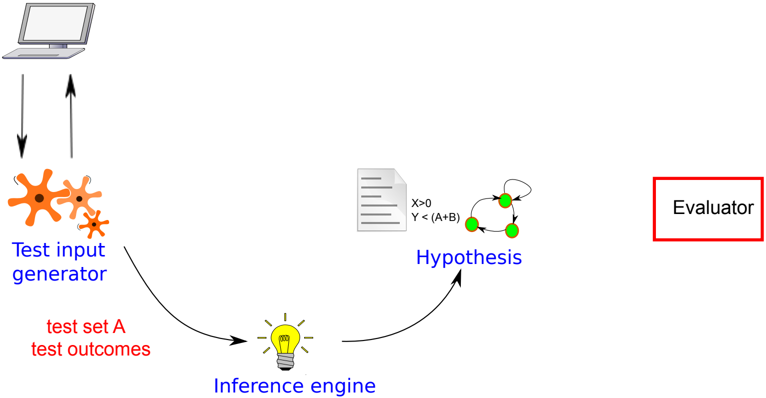
THE PROBABLY APPROXIMATELY CORRECT (PAC) FRAMEWORK



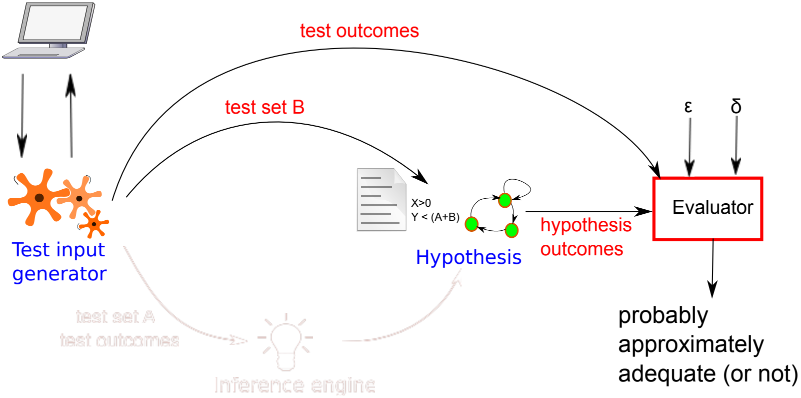
Using PAC to Assess Test Adequacy



Using PAC to Assess Test Adequacy



USING PAC TO ASSESS TEST ADEQUACY



Using PAC to Assess Test Adequacy

Assumptions

- Validity of final outcome must be interpreted with care
  - Test set is being evaluated against itself
  - Size of sets A and B must be sufficiently large and distinct
  - Test set generator must be capable of (eventually) exhaustively exercising the SUT

CONCLUSIONS

- Inferring models from tests gives us a 'test-eye view' of the system
- Test adequacy can be assessed by measuring model accuracy
- This can be achieved with established ML techniques
- For a given type of system (e.g. state-based) the PAC approach can be used to assess and compare the general performance of testing techniques.

Challenge

Find the best combination of machine-learner and test-set generator.