



INSTITUTO DE CIÊNCIAS EXATAS

DEPARTAMENTO DE COMPUTAÇÃO

Geração de Dados de Teste

Eduardo Kinder Almentero
ekalmentero@gmail.com

Conceitos

- **Domínio de entrada**

- Conjunto de todos os valores possíveis que podem ser utilizados para executar o programa.
- **Dado de teste** é um elemento do domínio de entrada.

- **Domínio de saída**

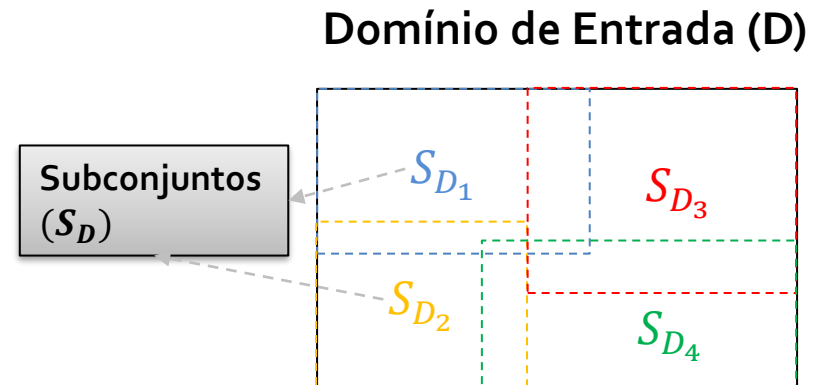
- Conjunto de todos os possíveis resultados que podem ser produzidos pelo programa.

- **Caso de teste**

- Par formado por um dado de teste e o comportamento esperado do software.

Geração de Dados de Teste

- Diante da **impossibilidade de executar um software com todos os seus valores possíveis de entrada**, a **seleção dos casos de teste com maior probabilidade de encontrar erros** é muito importante;
 - Para auxiliar nesta seleção, estudamos os **critérios de teste**, que basicamente, **dividem o domínio de entrada em subdomínios** (não necessariamente disjuntos).
- **Característica dos subconjuntos:**
 - Se um dos elementos do subconjunto **permite a descoberta de um erro**, os demais também permitirão;
 - De forma similar, se um elemento do subconjunto **não permite** encontrar um erro, os demais também não permitirão.
- Neste caso, o teste requer a execução de **pelo menos um elemento de cada subconjunto**;



Geração de Dados de Teste

- A geração de caso de testes é uma **tarefa complexa**, envolve um **grau de incerteza**, e diversos **fatores precisam ser considerados**:
 - Facilidade de automatização;
 - Redução de esforço do testador;
 - Requisitos funcionais e não funcionais
 - Custo, tempo de execução, uso de memória e cobertura de elementos necessário por um critério de teste.
- Não há um **algoritmo de propósito geral** para **criar um conjunto de casos de teste** que **satisfaça um critério**;
 - Também **não é possível** determinar **se esse conjunto existe!**
- O **problema** de geração de casos de teste é **indecidível**;
- **Restrições** relacionadas a **natureza da atividade de teste** **impossibilitam** que a **etapa de geração de casos de teste** seja **totalmente automatizada**;

Geração de Dados de Teste

- **Problemas que dificultam a automação da geração e casos de teste:**
 - **Correção coincidente**
 - Um caso de teste alcança um defeito do software, um estado de erro é produzido, mas coincidentemente o resultado correto é obtido;
 - **Caminho ausente**
 - O caminho corresponde a uma funcionalidade do software não existe (por algum motivo não foi implementado);
 - **Caminhos não executáveis**
 - Não existe um valor de entrada do software que levem à execução de um caminho;
 - **Mutantes equivalentes**
 - A modificação do programa pode não alterar o comportamento da funcionalidade implementada;

Geração de Dados de Teste

- As diferentes **técnicas de geração de dados** são divididas de acordo com suas características:
 - Aleatória ou direcionada
 - Totalmente ao acaso ou há alguma diretriz.
 - Estática ou dinâmica
 - Se envolve ou não a execução de software.
 - Otimização exata
 - Envolvendo programação linear.
 - Baseada em busca.
 - Por exemplo utilizando algoritmos genéticos.
- Algumas técnicas são **mais adequadas a tipos específicos de teste**
 - Teste de estresse,
 - Teste de requisitos não funcionais.

Teste Aleatório (TA)

- Na geração aleatória, também conhecida como **Teste Aleatório (TA)**, os **dados são selecionados de forma aleatória** do domínio de entrada, a partir de um gerador pseudoaleatório;
- Esta **seleção** é feita de acordo com alguma **distribuição estatística**, como a **distribuição uniforme**, onde cada elemento do domínio de entrada tem **chance igual de ser selecionado**;
- O TA também pode considerar **perfil real**, que considera **como o software será utilizado no contexto real**, e **pondera as entradas** de acordo com as **probabilidades** destas ocorrerem durante o uso real;

Teste Aleatório (TA)

- O **TA** é frequentemente **comparado** com as **técnicas baseadas em particionamento**;
- Algumas **análises empíricas** realizadas (olhar livro) tem tido **resultado positivo para TA**;
 - Considerando **custo-benefício** e **eficácia**
- Abordagens baseadas em TA se distinguem através da **forma como os dados são utilizados na execução do software**
- Neste contexto, **duas abordagens são mais comuns**:
 1. Utilizar os dados aleatoriamente como **um conjunto de dados de teste** para executar o software e **analisar o comportamento resultante em relação ao esperado**;
 2. A abordagem aleatória é utilizada para **gerar dados de teste para critérios baseados em particionamento**.

Teste Aleatório Adaptativo (Adaptative Random Testing (ART))

- Esta abordagem utiliza **dois conjuntos** de casos de teste: o **conjunto de testes executados (CTE)** e o **conjunto de testes candidatos (CTC)**;
- O CTC contém um **conjunto de dados de teste gerado aleatoriamente**;
- O CTE é atualizado com elementos selecionados do CTC;
- A cada incremento do CTE, **um dado de teste do CTC é selecionado** de acordo com **as distâncias dos dados de teste do CTC para os dados de teste do CTE**;
 - É selecionado o dado de teste que possui o maior valor de distância mínima;
 - Para o cálculo destas distâncias é utilizada a distância euclidiana entre dados;
 - Outra opção é a distância de Levenshtein para strings;

Considerações

- **Geração aleatória de dados**
 - tem como **vantagens** o **baixo custo para execução** e a **facilidade de automatização**;
 - a **principal desvantagem** é que ela **não considera a execução dos dados relacionados aos critérios de teste**, **nem promove a seleção dos elementos mais relevantes** do domínio de entrada;
 - Outra desvantagem é que **não agrega conhecimento sobre o software** e **não ajuda na identificação de elementos não executáveis e mutantes equivalentes**.
- **Extensões, como a ART, buscam lidar com estas desvantagens;**
- **A geração de dados de teste estática**, com base em restrições, tem evoluído mecanismos mais sofisticados;
 - A vantagem destas técnicas é que **não exigem a execução do software**.

Dados Realísticos

- Simulam com **maior precisão** as **condições reais** de **uso** do software;
 - Importante, principalmente no contexto de SI, para estimular o teste.
- **Auxiliam na comunicação**
 - Demonstração de funcionalidades do software
- Contém **características** que **não foram antevistas**
 - Por exemplo, texto com diversos estilos de escrita e pontuação;
- Permitem encontrar **lacunas** até mesmo na **especificação dos requisitos**

Dados Realísticos

- Como criar dados realísticos?
 - É preciso considerar a regionalização;
 - Plataformas para geração de dados
 - Mockaroo
 - Generatedata.com
 - Random Data Generator
 - Randon Data API
 - Fake.js
 - ...
 - Expressões Regulares (Regex)
 - Customização de dados



INSTITUTO DE CIÊNCIAS EXATAS

DEPARTAMENTO DE COMPUTAÇÃO

Perguntas?