

Automatização de oráculos de teste para imagens médicas de modelos tridimensionais

Misael C. Júnior¹, Márcio E. Delamaro¹

¹Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo (USP)
Avenida Trabalhador São-Carlense, 400 - Centro CEP: 13566-590 – São Carlos – SP

`misaeljr@usp.br`, `delamaro@icmc.usp.br`

Resumo. *Oráculos de teste determinam se uma execução de um SUT (do inglês, System Under Test) está correta ou não. Entretanto, dependendo da natureza dos dados produzidos pelo sistema, o SUT é conhecido como sistema de saída complexa, tornando a automatização dos oráculos um desafio. Sistemas na área de saúde, em particular, que analisam imagens tridimensionais, exemplificam um tipo de sistema de saída complexa. Um dos desafios associados a sistemas que analisam imagens tridimensionais é saber se a saída produzida está correta ou não. O fato de se tratar de um sistema de saída complexa torna essa tarefa mais difícil, fazendo com que estratégias ad-hoc e manuais sejam aplicadas. Nesse trabalho de mestrado procurou-se contribuir por meio da definição de oráculos de teste baseados na extração de características das saídas do sistema. A abordagem proposta foi aplicada especificamente em sistemas cujas saídas consistem em imagens sintéticas tridimensionais de vasos sanguíneos. Para tanto, é explorado o framework O-FIm/CO (do inglês, Oracle for Images and Complex Outputs), que utiliza conceitos de CBIR (do inglês, Content-Based Image Retrieval) como uma forma de automatizar oráculos de teste. Além de adaptações e extensões do framework, desenvolveram-se plug-ins, que representam extratores de características para imagens sintéticas tridimensionais de vasos sanguíneos. Dois estudos experimentais foram conduzidos objetivando avaliar a eficácia e a precisão dos oráculos de teste baseados em características na avaliação desse tipo de imagem. Além disso, realizou-se um estudo experimental comparando oráculos automatizados e oráculos humanos. Os resultados evidenciam a eficácia da abordagem como uma estratégia promissora para automatizar atividades de teste, contribuindo para a redução de tempo e esforços gerados por abordagens manuais durante a avaliação da qualidade de sistemas geradores de imagens médicas tridimensionais.*

1. Caracterização do problema e Motivação

O teste de software é um dos mais importantes componentes da área de Engenharia de Software (ES). Conceitualmente, teste de software é o conjunto de atividades dinâmicas que consistem na execução de um programa com algumas entradas específicas, visando verificar se o seu comportamento é condizente com sua especificação [Bertolino 2007, Delamaro et al. 2017]. Dentro do contexto de teste, oráculos de teste são os mecanismos utilizados para julgar a correção da saída ou o comportamento de uma execução de um determinado programa [Oliveira et al. 2009].

Um problema recorrente associado à automatização de oráculos de teste ocorre quando a saída do SUT (do inglês, *System Under Test*) configura um domínio complexo de dados [Oliveira et al. 2014]. Nesse cenário, a alta complexidade das saídas produzidas pelo SUT dificulta a aplicação das atividades de teste [Oliveira et al. 2014]. Atualmente, muitos são os exemplos de sistemas de saídas complexas como: (1) sistemas com GUIs

(do inglês, *Graphical User Interface*) complexas; (2) sistemas TTS (do inglês, *Text-To-Speech*); (3) aplicações Web; (4) aplicativos móveis; (5) ambientes de Realidade Virtual (RV); e (6) sistemas que analisam imagens médicas tridimensionais (3D).

Em sistemas biomédicos – exemplo de sistema cujas saídas são consideradas complexas – é essencial a aplicação de atividades que objetivem avaliar sua qualidade. Sistemas biomédicos são aqueles que têm a capacidade de apoiar a decisão clínica de profissionais da área médica. Sistemas que manipulem imagens sintéticas 3D de vasos sanguíneos exemplificam um tipo de sistema biomédico, contribuindo na antecipação do diagnóstico de determinadas patologias [Galarreta-Valverde et al. 2013]. No entanto, sistemas que manipulam imagens sintéticas 3D de vasos sanguíneos não têm sua qualidade mensurada devido à falta de estratégias de teste automatizado.

Em decorrência da falta de estratégias que visem automatizar o processo de avaliação de qualidade de tais sistemas, estratégias *ad-hoc* e manuais são aplicadas. Uma estratégia promissora para avaliar a qualidade de sistemas que manipulem tais imagens é por meio de uma abordagem denominada “oráculos de teste baseados em características” [Delamaro et al. 2013, Oliveira 2017]. Oráculos de teste baseados em características é uma abordagem na qual as características extraídas das saídas do SUT são utilizadas como fonte de informação para o oráculo julgar a correção da saída do SUT.

No presente trabalho de mestrado procurou-se definir e avaliar a abordagem “oráculos de teste baseados em características” para dar suporte a atividades de avaliação de sistemas que geram ou processam imagens sintéticas 3D de vasos sanguíneos. Devido a problemas recorrentes de falta de qualidade em tais sistemas, em consequência da falta de estratégias de teste automatizado durante o seu processo de desenvolvimento, considera-se necessária aplicações específicas para o suporte adequado ao teste de sistemas dessa natureza [Galarreta-Valverde et al. 2013].

Para tanto, a abordagem proposta explora conceitos de CBIR (do inglês, *Content-Based Image Retrieval*) para promover a automatização de oráculos de teste para domínios de saídas complexas. CBIR é uma técnica oriunda da área de Processamento de Imagem e Reconhecimento de Padrões que permite que um sistema recupere um conjunto de imagens a partir de um banco de imagens, tendo como chave de busca uma imagem de referência. Nesse cenário, o *framework* O-FIm/CO (do inglês, *Oracle For Images and Complex Outputs*) [Oliveira et al. 2009, Delamaro et al. 2013] emprega conceitos de CBIR para permitir que sejam criados oráculos que apoiem o teste automatizado em sistemas com saídas complexas como, no presente caso, sistemas de geração de imagens 3D de vasos sanguíneos.

Desse modo, dois principais fatores motivaram o desenvolvimento do presente trabalho: (i) estratégias tradicionais de teste são ineficazes quando aplicadas em sistemas com saída complexa; e (ii) estratégias sistemáticas e/ou automatizadas de teste configuram como um novo campo de pesquisa em sistemas biomédicos. Esse artigo sumariza os principais resultados e contribuições obtidas por meio do presente trabalho de mestrado. Para maiores detalhes, recomenda-se verificar o documento de dissertação disponível na Biblioteca Digital de Teses e Dissertações da Universidade de São Paulo (USP)¹.

2. Objetivos e Contribuições

O objetivo geral do trabalho desenvolvido é o fornecimento de alternativas para uma avaliação criteriosa, sistemática e produtiva da qualidade de sistemas com saídas que

¹<https://goo.gl/MPRt3e>

processam ou geram imagens sintéticas 3D de vasos sanguíneos. Para tanto, utilizou-se a abordagem “oráculos de teste baseados em características”. Dado tal objetivo geral, definiram-se os seguintes objetivos específicos:

- definir o modo como extratores de características podem fornecer informações sobre a correção de sistemas que processem ou gerem imagens sintéticas 3D de vasos sanguíneos;
- documentar e generalizar características e particularidades de sistemas que processem ou gerem imagens sintéticas 3D de vasos sanguíneos que possam ser utilizados em atividades de teste de software;
- generalizar um *framework* para teste de sistemas de saídas complexas; e
- estabelecer um ambiente para que novos sistemas 3D sejam inseridos no contexto do O-FIm/CO.

A partir dos objetivos traçados e descritos acima, podemos listar, em resumo, algumas contribuições em decorrência da execução do presente trabalho:

1. *Aplicação da abordagem oráculos de teste baseados em características em um novo domínio de sistemas com saídas complexas.*
A abordagem oráculos de teste baseados em características já havia sido aplicada, com sucesso, em outros domínios de sistemas complexos. Desse modo, o presente trabalho de mestrado contribui na definição, adaptação e aplicação da abordagem na avaliação de sistemas que geram ou processam imagens sintéticas 3D de vasos sanguíneos.
2. *Um Mapeamento Sistemático sobre abordagens e estratégias de teste de software e avaliação de qualidade em sistemas biomédicos.*
Durante a execução do presente trabalho, um Mapeamento Sistemático (MS) foi conduzido por meio da análise de 91 estudos publicados de 1990 a 2016. O MS objetivou identificar abordagens e estratégias para o teste de software e análise de qualidade em sistemas biomédicos, sintetizando a evidência disponível e identificando os principais problemas que envolvem avaliação de qualidade nesse domínio de pesquisa.
3. *Extensões na estrutura do framework O-FIm/CO, desenvolvimento e validação de extratores de características para imagens médicas 3D com seus código-fontes disponibilizados de modo open-source.*
Para aplicação da abordagem por meio do *framework* O-FIm/CO, foram realizadas adaptações na estrutura do O-FIm/CO e um catálogo de *plugins* no formato de extratores de características foi desenvolvido. Os extratores de características foram disponibilizados em um repositório online como contribuição do trabalho².
4. *Avaliações experimentais de oráculos de teste como uma estratégia sistemática para a avaliação de sistemas que geram ou processam imagens sintéticas 3D de vasos sanguíneos, destacando suas vantagens e desvantagens.*
O primeiro estudo experimental, nomeado “Estudo de Configuração”, avalia a precisão da abordagem na categorização de imagens sintéticas 3D e 2D de vasos sanguíneos. Esse experimento foi essencial para estabelecer um escopo da utilização da abordagem proposta na avaliação de imagens médicas 3D, gerando subsídios fundamentais para condução de estudos futuros. O segundo estudo experimental, nomeado “Estudo de Avaliação”, investiga a precisão da abordagem na avaliação de imagens sintéticas 3D de vasos sanguíneos após aplicação de ruído Gaussiano. Por fim, o terceiro estudo experimental, nomeado “Estudo de Comparação”, visa complementar o “Estudo de Avaliação”, comparando precisão dos oráculos automatizados com oráculos humanos.

²<https://goo.gl/6cy4wt>

3. Resultados Obtidos

Inicialmente, as diretrizes necessárias para adaptar a abordagem “oráculos de teste baseados em características” para o domínio de sistemas que gerem e/ou processem imagens sintéticas 3D de vasos sanguíneos foram definidas. Em seguida, *plugins* no formato de extratores de características foram desenvolvidos para o *framework* O-FIm/CO. Por fim, os estudos experimentais foram planejados e conduzidos objetivando avaliar o desempenho da abordagem na avaliação de imagens sintéticas 3D de vasos sanguíneos. A Figura 1 apresenta uma visão geral do processo de aplicação da abordagem.

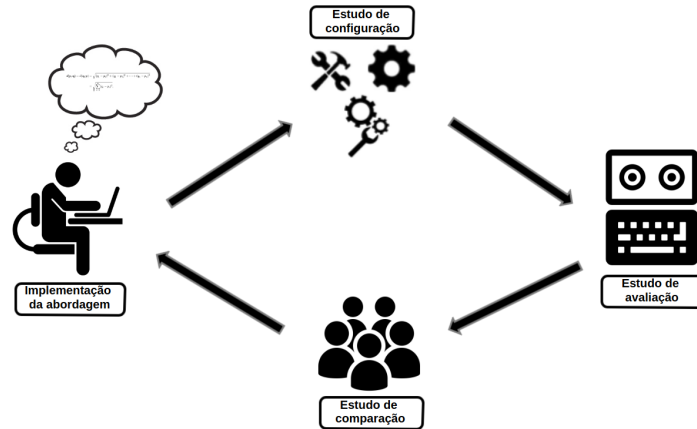


Figura 1. Visão geral do processo de aplicação da abordagem.

No “Estudo de Configuração”, uma imagem do conjunto de dados é considerada como modelo, enquanto as outras imagens do conjunto de dados são consideradas como imagens a serem avaliadas. As características de densidade, quantidade de pontos de bifurcação e terminais e quantidade de segmentos são extraídas da imagem modelo e das imagens a serem avaliadas. As imagens que o oráculo julga, por meio de uma métrica de similaridade, como similares à imagem modelo, são atribuídas a uma mesma categoria. Essa categorização do oráculo é comparada com a categorização manual realizada por dois especialistas em processamento de imagens médicas para avaliar o desempenho do oráculo. Por fim, um classificador randômico é utilizado para categorizar as imagens 3D e 2D e o desempenho do classificador randômico é comparado com o oráculo automatizado. A Figura 2 ilustra o processo de julgamento das imagens realizado pelo oráculo.

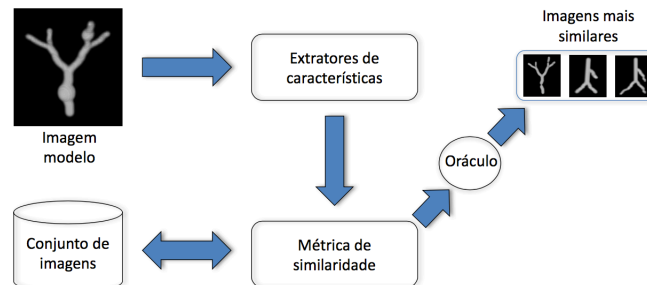


Figura 2. Processo de julgamento das imagens realizado pelo oráculo.

Na categorização das imagens 3D, o oráculo apresentou uma média de 77% de precisão, 100% de revocação e 88% de especificidade. Por outro lado, na categorização das imagens 2D, o oráculo apresentou uma média de precisão igual a 72%, 81% de revocação e 93% de especificidade. Em seguida, o teste binomial foi realizado para

verificar se havia uma diferença estatisticamente significativa em relação ao número de verdadeiros positivos (VP) e verdadeiros negativos (VN) dos oráculos automatizados. O teste binomial evidenciou que os oráculos automatizados atingiram uma alta significância estatística evidenciando seu alto desempenho em julgar as imagens corretamente.

Por fim, foi realizado o *t-test* com base no número de VP e VN dos oráculos em relação ao classificador randômico na categorização das imagens. Os testes realizados evidenciaram que o oráculo automatizado obteve resultados melhores que o classificador randômico e uma diferença considerada estatisticamente significativa (*p-value* = 0,0109). Tais resultados evidenciam que os oráculos tiveram um desempenho que lhes permite ser explorados em ambientes de teste.

No “Estudo de Avaliação”, conduzido exclusivamente com imagens 3D de vasos sanguíneos, inicialmente, cada imagem 3D é definida como modelo ou referência. Por outro lado, as imagens com ruído Gaussiano são definidas como imagens a serem avaliadas. As características de densidade, quantidade de pontos de bifurcação e terminais, quantidade de segmentos e tamanho médio dos segmentos são extraídas da imagem modelo e das demais imagens com ruído. Por fim, o oráculo automatizado, por meio de uma métrica de similaridade, julga se a imagem modelo é similar à respectiva imagem após aplicação de ruído Gaussiano. O processo de julgamento do oráculo automatizado é similar ao processo ilustrado na Figura 2.

O oráculo apresentou uma média de precisão igual a 100%, revocação igual a 85% e especificidade igual a 100%. Tais resultados evidenciam que o oráculo automatizado julgou corretamente as imagens a partir da imagem modelo, apesar de, em alguns casos, não ter retornado todas as imagens esperadas. Em seguida, o teste binomial foi realizado e evidenciou que os oráculos atingiram uma alta significância estatística (*p-value* = 1×10^{-12}), demonstrando sua alta aplicabilidade em avaliar tais imagens.

Finalmente, o “Estudo de Comparação” replicou o “Estudo de Avaliação” a partir do julgamento de 45 oráculos humanos. Para tanto, inicialmente, as imagens do “Estudo de Avaliação” foram inseridas, em um formulário de pesquisa juntamente com as informações necessárias para a realização do experimento. O *link* para acesso ao formulário de pesquisa foi distribuído e 45 oráculos humanos responderam ao formulário, avaliando a similaridade entre a imagem modelo e as respectivas imagens com ruído.

Os oráculos humanos apresentaram uma média de precisão igual a 73%, revocação igual a 72% e especificidade igual a 94%. Em seguida, o desempenho dos oráculos humanos foram comparados com o desempenho do oráculo automatizado por meio do *t-test* pareado. Com base na Tabela 1, o oráculo automatizado e os oráculos humanos são equivalentes em relação à métrica de revocação e estatisticamente não equivalentes em relação às métricas de precisão e especificidade. Com base na média de precisão e especificidade, o oráculo automatizado apresentou uma maior média de precisão e especificidade que os oráculos humanos, evidenciando a eficácia da abordagem na avaliação de sistemas dessa natureza.

Tabela 1. T-test entre o oráculo automatizado e os oráculos humanos.

Métrica	<i>p-value</i>	<i>t-value</i>	H ₀
Precisão	0.0018	4.8671	Rej.
Revocação	0.2000	1.4151	Ac.
Especificidade	0.0035	4.3245	Rej.

4. Discussão e Contribuições para a Literatura

O presente trabalho relata as experiências no uso da abordagem “oráculos de teste baseados em características” na avaliação de sistemas que geram ou processam imagens sintéticas 3D de vasos sanguíneos. Tais sistemas apresentam problemas recorrentes de qualidade devido à falta de estratégias de teste automatizado. Portanto, é essencial o uso de abordagens automatizadas para mensurar a qualidade de sistemas dessa natureza.

A partir dos resultados obtidos, a eficácia da abordagem pode ser destacada como uma estratégia promissora na avaliação de sistemas complexos, em específico, que analisam imagens médicas 3D. Assim, o presente trabalho apresentou três estudos experimentais que evidenciam a aplicabilidade de estratégias de teste em sistemas que processam imagens médicas 3D. Para tanto, configurações e extensões foram implementadas na arquitetura O-FIm/CO como suporte adequado para os oráculos de teste baseados em características. Além disso, um catálogo de *plug-ins* que representam extratores de características para imagens médicas 3D de vasos sanguíneos foi desenvolvido e disponibilizado de modo *open-source*.

Em relação às contribuições para a literatura, os resultados obtidos por meio do “Estudo de Configuração” foram reportados no 2nd Simpósio Brasileiro de Teste de Software Sistemático e Automatizado (SAST) em setembro de 2017³. Além disso, um artigo está em processo de revisão para submissão ao Journal Software Quality, visando reportar os resultados obtidos por meio deste trabalho de mestrado. Espera-se que os resultados apresentados possam motivar outros pesquisadores a se esforçarem para aplicar conceitos e estratégias de validação da ES em sistemas com saídas complexas.

Referências

- Bertolino, A. (2007). Software testing research: Achievements, challenges, dreams. In *Future of Software Engineering*, pages 85–103.
- Delamaro, M. E., Maldonado, J. C., and Jino, M. (2017). *Introdução ao teste de software*. Elsevier.
- Delamaro, M. E., Nunes, F. L. S., and Oliveira, R. A. P. (2013). Using concepts of content-based image retrieval to implement graphical testing oracles. *Software Testing, Verification and Reliability*, pages 171–198.
- Galarreta-Valverde, M. A., Macedo, M. M. G., Mekkaoui, C., and Jackowski, M. P. (2013). Three-dimensional synthetic blood vessel generation using stochastic L-systems. In *Proceedings of the Medical Imaging: Image Processing*, pages 86691I–86691I–6.
- Oliveira, R. A. P. (2017). *Test oracles for systems with complex outputs: the case of TTS systems*. Ph.d. thesis, Instituto de Ciências Matemáticas e de Computação (ICMC) – Universidade de São Paulo (USP), São Carlos, SP.
- Oliveira, R. A. P., Delamaro, M. E., and Nunes, F. L. S. (2009). O-FIm – oracle for images. In *Simpósio Brasileiro de Engenharia de Software (SBES)*, pages 1–6.
- Oliveira, R. A. P., Gil, V. N., Nunes, F. L. S., and Delamaro, M. E. (2014). An extensible framework to implement test oracle for “non-testable programs”. In *Proceedings of the 26th International Conference on Software Engineering & Knowledge Engineering*, pages 199–204.

³<https://goo.gl/yURMBp>