

ESTUDO SOBRE CONFIGURAÇÃO DE ALGORITMOS GENÉTICOS MULTIOBJETIVOS PARA OTIMIZAÇÃO DE PROJETO DE ARQUITETURA DE LINHA DE PRODUTO

Narcizo Gabriel Freitas Palioto (PIBIC/CNPq/FA/Uem), Thelma Elita Colanzi (Orientador), e-mail: narcizo.gabriel2@gmail.com.

Universidade Estadual de Maringá / Centro de Tecnologia/ Maringá, PR.

Ciência da Computação/ Engenharia de Software

Palavras-chave: otimização de parâmetros, algoritmo genético, arquitetura de linha de produto de software

Resumo:

Este trabalho envolveu um estudo sobre a melhor configuração de parâmetros do algoritmo NSGA-II, um algoritmo multiobjetivo baseado em algoritmo genético (AG), tendo em mente sua aplicação na ferramenta OPLA-Tool para otimização de arquitetura de Linhas de Produto de Software (LPS). Após uma revisão da literatura definiu-se por estudar empiricamente os melhores valores para os seguintes parâmetros de AGs: Tamanho da população, Número de Avaliações de *Fitness*, Funções Objetivos e Probabilidade de Mutação. Após a execução de 26 experimentos usando duas LPS, os melhores valores para cada parâmetro foram identificados usando o indicador de qualidade hypervolume e testes estatísticos.

Introdução

A abordagem de Linha de Produto de Software (LPS) é uma das principais alternativas para maximizar o reuso de software e, por isso, tem sido amplamente utilizada na indústria e abordada na academia (ARCURI e FRASER, 2011). O principal artefato dessa abordagem é a arquitetura de LPS (PLA – *Product Line Architecture*). Porém, a obtenção da PLA não é uma tarefa trivial devido à necessidade de atender a fatores relacionados a diferentes propriedades arquiteturais, os quais muitas vezes são conflitantes, por exemplo, modularidade de características, reusabilidade e extensibilidade da LPS (COLANZI, 2014).

Problemas da Engenharia de Software similares a este têm sido eficientemente resolvidos com algoritmos de busca em um campo de pesquisa conhecido como Engenharia de Software Baseada em Busca (SBSE - *Search Based Software Engineering*) (HARMAN e JONES, 2001). A SBSE converte um problema de engenharia de software num problema de busca computacional que pode ser resolvido com uma metaheurística.

Geralmente o problema demanda uma área muito grande para que uma busca exaustiva tenha um desempenho satisfatório, sugerindo uma aproximação usando metaheurísticas.

Combinando os conceitos de LPS e SBSE foi criada a MOA4PLA (*Multi-Objective Approach for Product-Line Architecture Design*) (COLANZI, 2014), uma abordagem que oferece um tratamento multiobjetivo para avaliar e otimizar projetos de PLA. Essa abordagem foi implementada na ferramenta OPLA-Tool utilizando o algoritmo NSGA-II (*Non-dominated Sorting Genetic Algorithm II*), um algoritmo multiobjetivo muito eficiente e popular para resolver problemas de otimização. O NSGA-II é uma extensão em algoritmos genéticos, por esse motivo ele usa operadores de seleção, cruzamento e mutação.

Para utilizar o NSGA-II é preciso configurar as probabilidades de aplicação dos operadores de mutação e de cruzamento, bem como qual o tamanho da população de indivíduos e o número de gerações que serão executadas ao longo do processo evolutivo. Neste contexto, foi realizado neste projeto um estudo a fim de descobrir qual a melhor configuração dos parâmetros do algoritmo NSGA-II especificamente para a aplicação na ferramenta OPLA-Tool a fim de otimizar projeto das PLAs Arcade Game Maker (AGM) e Mobile Media (MM) no contexto da MOA4PLA.

Materiais e métodos

Durante a execução do trabalho foi realizada uma revisão literária que teve como objetivo, além do aprofundamento nos temas do projeto, a definição de quais parâmetros seriam levados em consideração no estudo. Tomando com base a literatura, os parâmetros escolhidos foram: número de avaliações de *fitness*, tamanho da população, probabilidade de mutação e número de execuções (Colanzi, 2014; Guizzo, 2014; Arcuri e Fraser, 2011; Sayyad et.al, 2013).

Os valores experimentados para cada parâmetro foram escolhidos com base nos trabalhos de Colanzi (2014), Guizzo (2014), Arcuri e Fraser (2011) e Sayyad et.al (2013). Primeiramente temos 12 configurações para cada PLA sendo elas compostas pela permutação dos parâmetros de número de avaliação de *fitness* e tamanho da população e fixado os valores de 0.9 para probabilidade de mutação e 15 para o número de execuções. Após o término da experimentação com essas configurações a melhor parametrização encontrada foi testada novamente reajustando o valor de probabilidade de mutação para 0.8 e comparando o resultado obtido com o resultado previamente observado para poder encontrar o melhor valor de parametrização. Com isso totalizam 26 configurações para as duas PLAs contabilizando 13 configurações para cada PLA. Todos os valores utilizados para cada parâmetro podem ser observados na Tabela 1.

Para saber a efetividade das configurações propostas foi usado o indicador de qualidade hypervolume. O hypervolume essencialmente mede a

área de cobertura que uma fronteira de Pareto conhecida exerce sobre o espaço de objetivos O valor de hypervolume foi coletado após cada execução do algoritmo, como são 15 execuções por configuração foram analisados 15 valores de hypervolume por configuração. Um detalhe importante é que para efetuar o cálculo de hypervolume os dados de saída de cada execução foram normalizados, ou seja, o intervalo dos valores foi fixado entre 0 e 1, com o objetivo de evitar que dados com valores discrepantes interfiram no cálculo do indicador de qualidade escolhido.

Parâmetro	Valores Utilizados
PLAs testadas	AGM, MM
Número de Avaliações de <i>fitness</i>	3000, 30000, 300000
Tamanho da População	10, 50, 100, 200
Probabilidade de Mutação	0.8, 0.9
Número de Execuções	15

Tabela 1: Valores de parâmetros utilizados nas configurações dos experimentos

Após a realização do cálculo de hypervolume foi aferida a normalidade da distribuição dos dados de cada configuração utilizando o teste de Shapiro-Wilk e através dele foi descoberto que a maioria absoluta das configurações tem uma distribuição não normal fazendo assim com que o próximo teste que tem o objetivo de verificar a diferença estatística entre as configurações seja o teste de Kruskal-Wallis. Para a confecção desse teste foram criados quatro subgrupos de configurações, sendo que cada subgrupo contém o tamanho da população como atributo em comum. Depois disso o teste foi rodado novamente, porém dessa vez com a melhor configuração de cada subgrupo para, então finalmente, se obter a melhor configuração geral da PLA.

Resultados e Discussão

Após todos os testes e cálculos estatísticos foram encontrados os melhores valores para cada PLA como é mostrado na Tabela 2.

PLA	Número de avaliações de <i>fitness</i>	Tamanho da População	Probabilidade de Mutação
AGM	3000	200	0.9
MobileMedia	3000	200	0.9

Tabela 2: Melhores valores de configurações dos parâmetros

Pode se observar que os valores de ambas as configurações encontradas coincidem e que ambas contêm o menor valor de número de avaliações *fitness* das configurações propostas, sendo assim uma das configurações com o menor tempo de execução perdurando por aproximadamente 1 hora enquanto uma configuração com 300000 avaliações *fitness* por exemplo leva no mínimo 5 dias. Esses resultados diferem dos resultados encontrados nos trabalhos de Colanzi (2014), onde os parâmetros usados foram de 30000 avaliações *fitness*, 100 de população e 0.9 de probabilidade de mutação e de Guizzo (2014), onde os parâmetros não foram fixados para todas as execuções, mas pode-se destacar dentre as configurações obtidas os parâmetros de tamanho da população, onde o

valor varia de 50 para 200 dependendo da PLA envolvida, e número de avaliações *fitness* e probabilidade de mutação onde foram usados os valores 30000 e 0.9, respectivamente.

Conclusões

O objetivo desse trabalho foi descobrir empiricamente as melhores configurações de parâmetros do algoritmo NSGA-II no contexto da MOA4PLA e da ferramenta OPLA-Tool. Com base nos trabalhos da literatura foi elaborado um conjunto de configurações de parâmetros focando em duas PLAs e dentre esse conjunto foi encontrada a configuração que encontra os melhores resultados de acordo com o indicador de qualidade hypervolume. Os resultados encontrados permitirão a obtenção de melhores resultados em futuros experimentos de otimização de projetos de PLA.

Agradecimentos

Agradeço ao PIBIC/CNPq/FA/Uem pelo apoio financeiro.

Referências

Arcuri, A., Fraser, G. **On parameter tuning in search based software engineering**. In Symposium on Search Based Software Engineering, Szeged, Hungary, 2011, pp. 33-47.

Colanzi, T. E.. **Uma abordagem de otimização multiobjetivo para projeto arquitetural de linha de produto de software**. Tese de doutorado, Universidade Federal do Paraná, Curitiba, PR, 2014.

Guizzo, Giovanni. **Uso de padrões em projeto arquitetural baseado em busca de linha de produto de software**. 2014. 168 f. Tese (Mestrado em Ciência da Computação), Universidade Federal do Paraná, Curitiba, 2014.

Harman, Mark; Jones, Bryan F. Search-based software engineering. **Information and software Technology**, v. 43, n. 14, p. 833-839, 2001.

Sayyad, Abdel Salam et al. **On Parameter Tuning in Search Based Software Engineering: A Replicated Empirical Study**. 2013 3rd International Workshop On Replication In Empirical Software Engineering Research, [s.l.], p.1-7, out. 2013.