FEDA-NRP: Un algoritmo de estimación de distribuciones multivariado y de estructura fija para resolver el problema de la próxima versión multiobjetivo con interacciones entre requisitos

Víctor Pérez-Piqueras[*] victor.perezpiqueras@uclm.es
Pablo Bermejo[*] pablo.bermejo@uclm.es
José A. Gámez[*] jose.gamez@uclm.es

Abstract

El problema de la siguiente versión (Next Release Problem) implica seleccionar el subconjunto óptimo de requisitos para una versión de software con el objetivo de maximizar la satisfacción de los agentes interesados mientras se minimiza el coste de desarrollo, considerando las dependencias entre requisitos. En un trabajo publicado recientemente [6] y aquí resumido, se propone un enfoque de optimización biobjetivo con restricciones utilizando un Algoritmo de Estimación de Distribución (EDA) que incorpora conocimiento del dominio. Se utiliza una red Bayesiana con una estructura fija para modelar las dependencias, reduciendo la complejidad de búsqueda y asegurando soluciones válidas. Además, se han generado datasets sintéticos con estimaciones de costes basadas en proyectos tanto ágiles como clásicos para evaluar el enfoque. Los resultados muestran que el método propuesto supera a otros algoritmos de optimización multiobjetivo en datasets complejos.

1 Introducción

Los proyectos software siguen un ciclo de vida estructurado que incluye la obtención, selección, diseño, implementación y pruebas de un conjunto de requisitos. Para abordar la complejidad y el alto consumo de tiempo de estas fases, la Ingeniería del Software Basada en Búsqueda (SBSE) reformula problemas de software en forma de tareas de optimización, permitiendo el uso de metaheurísticas como los algoritmos genéticos, algoritmos de estimación de distribución (EDAs), etc.

Este trabajo se centra en el Problema de la Próxima Versión (NRP) [1], un problema clave en SBSE que optimiza la selección de requisitos para una versión de software, equilibrando costes, satisfacción de los interesados y dependencias. El NRP es relevante tanto en metodologías clásicas como ágiles, donde requiere reevaluaciones frecuentes. Tradicionalmente, el NRP se ha abordado con optimización multiobjetivo. La versión multiobjetivo del Problema de la Próxima Versión (MONRP) [7] se define mediante un conjunto de requisitos candidatos $R = \{r_1, \ldots, r_n\}$, solicitados por clientes $C = \{c_1, \ldots, c_m\}$. Cada requisito tiene un coste $E = \{e_1, \ldots, e_n\}$ y una importancia asignada por los clientes, almacenada en una matriz v_{ij}. La satisfacción total de un requisito r_j es $s_j = \sum_{i=1}^{m} w_i \times v_{ij}$, donde w_i representa la importancia del cliente c_i. El MONRP busca un subconjunto de requisitos $X \subseteq R$ que maximice la satisfacción de los clientes $S(X) = \sum_{j \in X} s_j$ y minimice el costo de desarrollo $E(X) = \sum_{j \in X} e_j$. Además, deben respetarse restricciones de dependencia, como la implicación entre requisitos ($r_i \Rightarrow r_j$).

[*]Departamento de Sistemas Informáticos (DSI, ESII), Grupo de Sistemas Inteligentes y Minería de Datos (I3A), Universidad de Castilla-La Mancha, Albacete, 02071, Spain

Preprint. Under review.

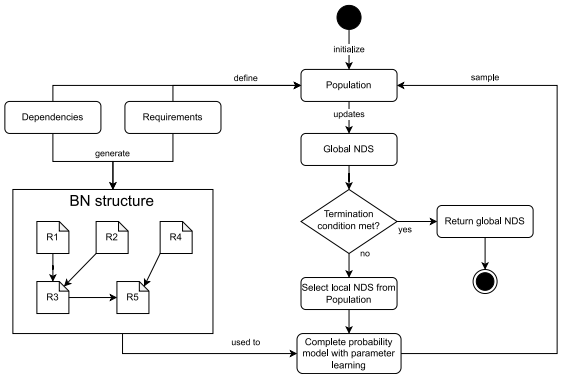

Figure 1: Diagrama de alto nivel de FEDA-NRP

En la literatura, el MONRP se ha resuelto frecuentemente mediante métodos metaheurísticos como ACO, PSO y NSGA-II [3, 2], pero los EDAs han sido poco explorados [5].

2 Propuesta

En [6] proponemos FEDA-NRP, un algoritmo de optimización biobjetivo basado en un EDA multivariado, diseñado para resolver el problema de la próxima versión multiobjetivo (MONRP). Este algoritmo aprovecha el conocimiento del dominio mediante una red Bayesiana de estructura fija que modela las dependencias entre requisitos utilizando un grafo acíclico dirigido (DAG). Esta estrategia evita el coste del aprendizaje estructural, lo que mejora la eficiencia y garantiza soluciones válidas. Además, se emplea un modelo probabilístico basado en puertas *leaky binary noisy-OR* [4], lo que reduce la complejidad espacial y de aprendizaje a la de un EDA univariado.

El proceso de FEDA-NRP sigue una estrategia PAES (Pareto Archived Evolutionary Strategy) y consta de varias etapas clave: inicialización del modelo probabilístico, muestreo de individuos según un orden ancestral, evaluación de soluciones y actualización del conjunto de soluciones no dominadas (NDS). La actualización de parámetros se realiza mediante estimación de máxima verosimilitud (MLE), manteniendo la probabilidad de selección para evitar convergencia prematura. El algoritmo mantiene dos archivos de soluciones NDS: uno global y otro local, usados para actualizar el modelo probabilístico en cada iteración. Su funcionamiento se ilustra en la Figura 1.

3 Resultados

La experimentación arranca con la creación de un corpus público de datasets con distintos niveles de complejidad y estimaciones de costes. Se compara FEDA contra tres EDAs (UMDA, PBIL y MIMIC) y dos algoritmos genéticos multiobjetivo propuestos en trabajos recientes (AGE-MOEA-II y C-TAEA). Utilizando sus mejores configuraciones de hiperparámetros, se ejecutan 30 veces en 14 conjuntos de datos. El proceso de evaluación se realiza sobre un conjunto de soluciones no dominadas ($NDS_{archive}$), seleccionado mediante una normalización min-max de los valores de satisfacción y coste, y usando un punto de referencia común para calcular el Hipervolumen (HV). Para una comparación justa, se seleccionan 10 soluciones del $NDS_{archive}$ de cada algoritmo mediante un proceso de selección greedy basado en HV. Además, se utilizan varios indicadores de calidad como HV, Distancia Generacional (GD+), Δ-Spread y Ratio de soluciones No Dominadas respecto al Frente de Referencia (UNFR), que permiten evaluar la convergencia, dispersión, uniformidad y contribución de cada algoritmo al frente no dominado de referencia. Los resultados se promedian sobre las 30 ejecuciones, asegurando la reproducibilidad de los experimentos mediante el uso de un entorno de ejecución consistente. El código y datos utilizados está disponible en un repositorio público[2], lo que garantiza la transparencia y replicabilidad del estudio.

En los experimentos realizados, FEDA-NRP mostró un rendimiento sobresaliente en términos de la métrica de HV (ver Figura 2), siendo el mejor algoritmo para conjuntos de datos con un

[2]https://github.com/UCLM-SIMD/MONRP/tree/eng_app_ai23

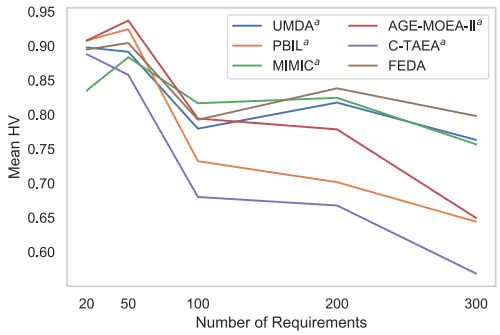

Figure 2: Hipervolumen medio en datasets con el mismo número de requisitos.

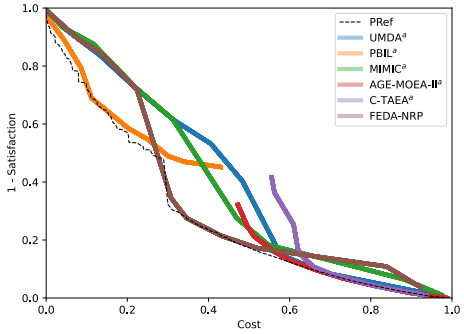

Figure 3: Indicador visual de las búsquedas de algoritmos para el conjunto de datos *d3*.

alto número de requisitos ($|R| \geq 200$), alcanzando el mayor HV en todos los casos, excepto en uno (a4, 200 requisitos, HV = 0.802 vs. 0.805 de MIMIC). En cuanto a las métricas de calidad, AGE-MOEA-II^a fue el mejor en términos de UNFR y GD+, mientras que UMDA^a destacó en Δ-Spread. Visualmente, FEDA-NRP fue capaz de explorar completamente el espacio de soluciones (ver Figura 3), encontrando soluciones cercanas al PRef (Frente de Pareto de Referencia) en zonas equilibradas y extremas, mientras que otros algoritmos como PBIL^a y MIMIC^a se centraron solo en un lado del espacio de Pareto (minimización de Coste o maximización de Satisfacción). Además, FEDA-NRP mostró un rendimiento significativamente mejor que C-TAEA^a, AGE-MOEA-II^a y MIMIC^a en términos de tiempo de ejecución, siendo aproximadamente 12 veces más rápido que C-TAEA^a y 5 veces más rápido que AGE-MOEA-II^a, lo que lo convierte en la opción más eficiente y efectiva para resolver el MONRP en proyectos con requisitos complejos.

Agradecimientos. Trabajo financiado por el Gobierno de Castilla-La Mancha, la UCLM y los Fondos Europeos, UE, mediante los proyectos SBPLY/21/180225/000062 y 2022-GRIN-34437.

Referencias

[1] Anthony J. Bagnall, Victor J. Rayward-Smith, and Ian M. Whittley. The next release problem. *Information & Software Technology*, 43(14):883–890, 2001.

[2] Jose M. Chaves-Gonzalez, Miguel Perez-Toledano, and Amparo Navasa. Software requirement optimization using a multiobjective swarm intelligence evolutionary algorithm. *Knowledge-Based Systems*, 83, 03 2015.

[3] José del Sagrado, Isabel del Águila, and FranciscoJ Orellana. Multi-objective ant colony optimization for requirements selection. *Empirical Software Engineering*, 20:577–610, 2015.

[4] Agnieszka Onisko, Marek J. Druzdzel, and Hanna Wasyluk. Learning Bayesian network parameters from small data sets: application of noisy-or gates. *International Journal of Approximate Reasoning*, 27(2):165–182, 2001.

[5] Víctor Pérez-Piqueras, Pablo Bermejo López, and José A. Gámez. Estimation of distribution algorithms applied to the next release problem. In *17th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2022)*, Lecture Notes in Networks and Systems, pages 98–108. Springer, 2023.

[6] Víctor Pérez-Piqueras, Pablo Bermejo, and José A. Gámez. Feda-nrp: A fixed-structure multivariate estimation of distribution algorithm to solve the multi-objective next release problem with requirements interactions. *Engineering Applications of Artificial Intelligence*, 124:106555, 2023.

[7] Yuanyuan Zhang, Mark Harman, and S Afshin. The multi-objective next release problem. In *Proceeding of the 9th annual conference on Genetic and Evolutionary Computation*, pages 1129–1137, 2007.