Staged Column Generation Approach for the Software Clustering Problem

Hugo Harry Kramer*†[1], Marcia Fampa[2], Viviane Köhler[3], Eduardo Uchoa[1], and François Vanderbeck[4,5]

[1]Universidade Federal Fluminense (UFF) – Brésil
[2]COPPE, Universidade Federal do Rio de Janeiro (COPPE/UFRJ) – Brésil
[3]Universidade Federal de Santa Maria (UFSM) – Brésil
[4]RealOpt (INRIA Bordeaux - Sud-Ouest) – Université Sciences et Technologies - Bordeaux I, INRIA – 200 avenue de la Vieille Tour 33405 Talence, France
[5]Institut de Mathématiques de Bordeaux (IMB) – CNRS : UMR5251, Université Sciences et Technologies - Bordeaux I, Université Victor Segalen - Bordeaux II, Université Sciences et Technologies Bordeaux I, Université Victor Segalen – Bordeaux II – 351 cours de la Libération 33405 TALENCE CEDEX, France

Résumé

The present work deals with the application of a novel Column Generation (CG) approach to the
Software Clustering Problem (SCP). In this problem, one has a Modular Dependence Graph (MDG)
consisting of a set V of nodes representing the modules of a software, a set E of edges, or the set
of relationships between two modules u, v ∈ V of the software, and each relationship has an associated
weight given by c_uv . The goal is to partition the MDG in clusters, and the quality of such partition
must be maximized. The quality of the partition is given by a measure called Modularization Quality
(MQ). To tackle this problem, we apply the Dantzig-Wolfe decomposition having a formulation from
literature as a starting point. Given this, we present two approaches based on CG: (i) the standard
CG approach, and (ii) a new approach, which we call Staged Column Generation (SCG). In this new
approach, at each stage, the problem is solved by CG considering a restricted version of the pricing
subproblem. At the last stage, the complete version of the subproblem is considered and CG runs
until convergence. In the context of the present work, in the last stage, the subproblems are solved by

*Intervenant
†Auteur correspondant: hugoharry@gmail.com

MIP, and in the previous stages, the restricted subproblems are solved heuristically. In addition, we
present a heuristic procedure to initialize the Master Problem, aiming to improve the convergence of
the algorithms. We test our algorithms in a set of 47 instances found in literature. With the proposed
approaches we were able to improve the literature results solving all these instances to optimality (45
of them in the root node). Furthermore, the SCG approach presented a considerable performance
improvement in terms of computational time, number of iterations and number of generated columns
when compared with the standard CG as the size of the instances grow.

Mots-Clés: Software Clustering Problem, Column Generation, Staged Column Generation