Project Team INRIA: RealOpt

- Authors of the proposal research subject: François Clautiaux and Pierre Pesneau

- Title of the proposal research subject: Reformulation and aggregation of mathematical models for clustering problems

- Scientific research context (10-15 lines):

Given a set of elements, clustering problems consist in affecting the elements to sets while minimizing a given objective function. A large quantity of research papers have been published on this topic. Classical clustering problems of medium to large size are now solved efficiently by state-of-the-art solvers. However, for very large scale problems (big data, biomedical research, ...), or problems with a more complex structure (constrained clusters, conflicts,...), the size of the mathematical models that have to be solved becomes too large.

From a methodological point of view, the last years have seen an increase of interest for aggregation methods in integer programming. Those methods solve smaller aggregated and thus approximate models and dynamically add information issued from disaggregation to the models until the optimality is reached. Complex clustering problems have specific structures that make them well suited for applying such aggregation methods.

- Goal (10-15 lines) :

The objective of the PhD thesis is to develop new aggregation strategies for solving efficiently several general clustering problems. The methodology we intend to develop is generic, and should be relevant for many different applications. However, we will test them against several real-life problems.
The goal is to propose innovative and efficient aggregation/disaggregation schemes in the context of extended formulations for clustering problems. In particular, we will consider several version of aggregation/disaggregation schemes for column generation algorithms: aggregation of master problem only, or aggregation of the master problem and subproblems. The goal of the thesis is also to determine the best way of computing an initial aggregation, and how to relate the different elements in the same aggregated element (equivalence, or restrictions on the possible dual values). In order to validate numerically their interest, we will test them against problem instances taken from real-life problems: first, on geographic clustering problems from an industrial partner, then on data from problems related to computer services and cloud computing.

- Project (10-15 lines) :

To carry out this research, the different steps of the project are the following:
1) Bibliography: there exists a wide bibliography on clustering problems as well as on aggregation/disaggregation methods. The PhD student will have to extract relevant papers dealing with mathematical programming and how very large scale clustering problems are handled until now.
2) Theoretical study of aggregation problems: comparison of different aggregation schemes, different disaggregation rules, and their impact on the pricing subproblems, methods for computing an initial aggregation

3) Application to the geographic clustering problem, practical comparison of different aggregation schemes
4) Application to the computer service problem
5) Embedding the algorithms into the optimization library of the team: the developed methods will be integrated to BaPCod (A Branch-and-Price Code), a library maintained by realopt team designed to help the development of column generation codes.

•Required Knowledge and background

Linear and integer programming, object-oriented programming

•Advisors : François Clautiaux and Pierre Pesneau

•References (max 5 lines) :

I. Elhallaoui, A. Metrane, F. Soumis, and G. Desaulniers, Multi-phase dynamic constraint aggregation for set partitioning type problems, Mathematical Programming 123 (2010), no. 2, 345-370.

I. Elhallaoui, D. Villeneuve, F. Soumis, and G. Desaulniers, Dynamic aggregation of set-partitioning constraints in column generation, Operations Research 53 (2005), no. 4, 632–645.

R. Macedo, C. Alves, J. Vaĺ rio de Carvalho, F. Clautiaux, and S. Hanafi, Solving exactly the vehicle routing problem with time windows and multiple routes using a pseudo-polynomial model, to appear in European Journal of Operational Research (2011).

R. Xu, and D. C. Wunsch. Clustering algorithms in biomedical research: A review. IEEE Reviews in Biomedical Engineering, 3, 120-154 (2010).

P. Hansen, and B. Jaumard. Cluster analysis and mathematical programming. Mathematical Programming, Series B, 79(1-3), 191-215 (1997).

•Keywords : clustering, mathematical programming, decomposition methods, aggregation methods, polyhedral analysis

•Duration : 3 years