Available online at www.sciencedirect.com
ScienceDirect
European Journal of Operational Research 192 (2009) 396–413
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/ejor

Discrete Optimization

POPMUSIC for the point feature label placement problem

Adriana C.F. Alvim [a], Éric D. Taillard [b,*]

[a] Department of Applied Informatics, Federal University of the State of Rio de Janeiro, Avenida Pasteur 458, Rio de Janeiro, RJ 22.290-240, Brazil
[b] Department of Industrial Systems, University of Applied Sciences of Western Switzerland, Route de Cheseaux 1, Case postale, CH-1401 Yverdon, Switzerland

Received 29 August 2006; accepted 3 October 2007
Available online 10 October 2007

Abstract

Point feature label placement is the problem of placing text labels adjacent to point features on a map so as to maximize legibility. The goal is to choose positions for the labels that do not give rise to label overlaps and that minimize obscuration of features. A practical goal is to minimize the number of overlaps while considering cartographic preferences. This article proposes a new heuristic for solving the point feature label placement problem based on the application of the POPMUSIC frame. Computational experiments show that the proposed heuristic outperformed other recent metaheuristics approaches in the literature. Experiments with problem instances involving up to 10 million points show that the computational time of the proposed heuristic increases almost linearly with the problem size. New problem instances based on real data with more than 13,000 labels are proposed.

Keywords: Metaheuristics; POPMUSIC; Map labeling; Large-scale optimization; Tabu search

1. Introduction

Automated label placement is a problem of fundamental importance in cartography, where text labels must be placed on maps while avoiding overlaps with cartographic symbols and other labels. Interactive creation of maps increases the importance of this problem. This task must be performed with limited computational effort, typically less than one second. Applications in cartography require different label placement tasks.

First, the object to be labeled may have several different dimensions:

- Dimension 0, labeling point features (such as cities and mountain peaks)
- Dimension 1, labeling line (segment) features (such as rivers and roads) and
- Dimension 2, labeling area features (such as countries and oceans)

Then, overlapping labels may be accepted or not. If two or more labels cannot overlap, two different problems can be defined: In the *label number maximization problem* [10], certain features (and their labels) are allowed to be deleted and the objective is to place as many labels as possible with no overlaps. This problem is equivalent to finding a maximum vertex independent set in a conflict graph [13,17] where each node is a candidate label and there is an edge between two nodes whenever there is a conflict between the corresponding labels. In the *label size maximization problem*, the objective is to determine the maximum scale factor for the label size and a corresponding labeling without overlaps.

If all features must be labeled and scaling is not allowed, overlaps must be permitted. In that case, there are two objectives: to minimize the number of overlaps, which is called the *label overlap minimization problem* by [10] and to minimize the number of labels obstructed by at least

* Corresponding author.
E-mail addresses: adriana@unriotec.br (A.C.F. Alvim), eric.taillard@heig-vd.ch (É.D. Taillard).

0377-2217/$ - see front matter
doi:10.1016/j.ejor.2007.10.002

one other label [3] which is called *maximum number of conflict free labels problem* by [12].

Concerning the position of the labels, there are two models. In the first one, an explicit enumerated set is considered for potential label positions (discrete model). In the second model, an infinite number of possible label positions may be used. This continuous model is also called the slider model.

Finally, for all labeling problems, it is possible to assign a weight (penalty) for each label position and to use as alternate objective the minimization of the sum of the penalties.

Construction of good labeling, regardless of the features being labeled, leads to combinatorial optimization problems that are generally NP-hard [7,9,11]. Exact algorithms are able to solve problems with just a few hundred points to label [5,10,13,22]. Therefore, heuristic algorithms must be designed for dealing with larger problems or for getting approximate solutions with low computational effort.

Although the methodology proposed in this paper can be applied to various labeling problems, we are going to illustrate this methodology on the NP-Hard point-feature label placement problem (PFLP) with *label overlap minimization*. The size of the labels is fixed, the potential positions of labels are discrete and all points must be labeled. Therefore, the first objective considered is to minimize the number of overlaps. Cartographic preferences can also be considered as an alternate objective.

Christensen et al. [4] presented a good review on the PFLP. The authors developed a local search technique based on a discrete form of gradient descent and a simulated annealing based algorithm. Verner et al. [16] proposed a heuristic based on genetic algorithms (GA). More recent works, considering that all point features must be labeled, include a tabu search [19], a constructive genetic approach [21] and a fast algorithm for label placement [20]. Wagner et al. [17] proposed a two phase algorithm for the *label number maximization problem*. An extensive Map-Labeling bibliography is maintained by Wolff [18].

This work, outlined in Alvim and Taillard [1], is based on the preliminary work of Burri and Taillard [2,14] which investigates the evaluation of the POPMUSIC methodology for the PFLP. POPMUSIC is a *general* optimization method especially designed for optimizing large instances of combinatorial problems and can be seen as a large scale neighborhood search (see [15]). The basic idea of POPMUSIC is to locally optimize sub-parts of a solution, once a solution of the problem is available. These local optimizations are repeated until no improvements are found. The local optimizations are performed by a new implementation of the tabu search proposed by [19].

Although the results presented in the present article are restricted to this labeling problem, this methodology can be used to deal with other labeling problems due to the ability of the tabu search used as local optimizer in POPMUSIC.

The paper is organized as follows: Section 2 introduces the PFLP problem. Then, the adaptation of POPMUSIC for the PFLP is presented in Section 3. This section starts by describing the general POPMUSIC frame (Section 3.1). A practical implementation based on POPMUSIC requires the design of few components specific to the problem being solved. The first one is the way an initial solution is obtained (Section 3.2.1), then the way subproblems are built (Section 3.2.2) and finally a procedure for optimizing the subproblems (Section 3.2.3). The last is based on a tabu search proposed by [19]. The tabu search is presented in more detail than in [19] (and is certainly slightly different), making the present article self-contained. Computational results and new instances are presented in Sections 4 and 5. Concluding remarks are made in the last section.

2. The point feature label placement problem

In this paper, we consider a set of n points that have to be labeled. Each point has p candidate label positions of identical size, identified by the integers $1,\dots,np$. The position of the label associated to point x, $x = (1,2,\dots,n)$, is given by variable y_x that can take p different values: $y_x \in \{(x-1)*p+1,(x-1)*p+2,(x-1)*p+3,\dots,x*p\}$.

Fig. 1 shows the possible label positions for a point feature when $p = 4$. Each box corresponds to a region in which the label may be placed. According to cartographic standards, there is a preference (or, more precisely, a penalty) for each possible label position, lower values indicating better positions: top right (position 1), top left (position 2), bottom left (position 3) and bottom right (position 4). An arbitrary weight $w(y_x) < 1$ is associated with each label position y_x. In this paper, we have considered problem instances with $p \in \{2,4,8\}$ label positions for each point x and respective weights $w(y_x) = ((y_x - 1) \bmod p) * 0.0001$. We are also given an overlap symmetrical $np \times np$ matrix A where $a_{ij} = 1.0 + w(j)$ if label i overlaps with label j, $a_{ij} = w(i)$ for $i = j$ and $a_{ij} = 0$ otherwise. A solution S is a list of n labels $(y_1, y_2, \dots, y_n)$. For a given solution S, the cost measure which counts the number of point features labeled with one or more overlaps is expressed by $f(S) = \sum_{i=1}^{n} \min\{1, \sum_{j\in\{1,\dots,n\}\setminus i} a_{y_i y_j}\}$; and the function that counts the number of overlaps and takes the cartographic preferences into account is expressed by $\bar{c}(S) = \sum_{i=1}^{n}\sum_{j=1}^{n}(a_{y_i y_j})$. For the special case where $w(y_x) = 0$, for $x = 1,\dots,n$, we note by $c(S)$ the function which simply

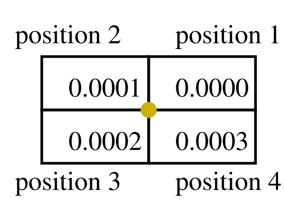

Fig. 1. A point feature (circle) with four potential label positions (boxes). A weight (penalty) $\{0, 0.0001, 0.0002, 0.0003\}$ is associated with each label position, lower values indicating best positions according to cartographic standards.