Expert Systems with Applications 42 (2015) 4432–4442

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Metaheuristics for large-scale instances of the linear ordering problem

Celso S. Sakuraba [a], Débora P. Ronconi [a,*], Ernesto G. Birgin [b], Mutsunori Yagiura [c]

[a] Production Engineering Department, Polytechnic School, University of São Paulo, Av. Prof. Almeida Prado, Trav. 2, No. 128 – Cidade Universitária, São Paulo 05508-070, Brazil
[b] Department of Computer Science, Institute of Mathematics and Statistics, University of São Paulo, Rua do Matão, 1010, Cidade Universitária, São Paulo 05508-090, Brazil
[c] Department of Computer Science and Mathematical Informatics, Graduate School of Information Science, Nagoya University, Furocho, Chikusaku, Nagoya 464-8603, Japan

ARTICLE INFO

Article history:
Available online 31 January 2015

Keywords:
Metaheuristics
Iterated local search
Great deluge
Linear ordering problem
Large-scale instances

ABSTRACT

This paper presents iterated local search and great deluge trajectory metaheuristics for the linear ordering problem (LOP). Both metaheuristics are based on the TREE local search method introduced in Sakuraba and Yagiura (2010) that is the only method ever applied to a set of large-sized instances that are in line with the scale of nowadays real applications. By providing diversification and intensification features, the introduced methods improve all best known solutions of the large-sized instances set. Extensive numerical experiments show that the introduced methods are capable of tackling sparse and dense large-scale instances with up to 8000 vertices and 31,996,000 edges in a reasonable amount of time; while they also performs well in practice when compared with other state-of-the-art methods in a benchmark with small and medium-scale instances.

1. Introduction

The linear ordering problem (LOP) was first described by Chenery and Watanabe (1958) in the context of the economic input–output studies proposed by Leontief (1966). By using input–output matrices that represent the flow among production sectors of a certain region, it is possible to analyze the stability of the regions' economy. In order to make this analysis, the order of the rows and columns of the matrix must be rearranged in a way that maximizes the sum of the values above the diagonal of the matrix, where an identical permutation must be used to rearrange the rows and the columns. Then the sectors that correspond to bigger suppliers and consumers can be identified. The LOP can also be seen as a graph problem, as shown in the example of Fig. 1. In the graph of Fig. 1(b), each vertex v_i corresponds to the i-th row and column of the matrix of Fig. 1(a), and the forward edges, represented above the vertices, correspond to the values in the upper triangle of the matrix. The equivalence of the two representations is immediate, e.g. by regarding the matrix as the adjacency matrix of the graph.

We formally define the LOP in its graph representation as follows: given a directed graph $G = (V, E)$ with vertices set V ($|V| = n$), edges set $E \subseteq V \times V$ ($|E| = m$), and a cost c_{uv} for each edge $(u, v) \in E$, find a permutation of the vertices that maximizes the total cost of the direct edges, i.e. edges directed from a vertex u to a vertex v with v being a vertex in a position after u in the permutation. The sum of the direct edges corresponds to the sum of the values above the diagonal of the matrix. We assume without loss of generality that $c_{uv} > 0$ holds for all $(u, v) \in E$ and that if we regard G as an undirected graph, it is connected (which implies $m \geqslant n - 1$). For convenience, we also assume $c_{uv} = 0$ for all $(u, v) \notin E$. Denoting a permutation by $\pi : \{1, \ldots, n\} \to V$, where $\pi(i) = v$ means that v is the i-th element of π, the total cost of the direct edges is formally defined as follows:

$$cost(\pi) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} c_{\pi(i)\pi(j)}. \quad (1)$$

Maximizing this sum is equivalent to minimizing the sum of the reverse edges (edges of the form (u, v) such that v appears before u in the permutation) or minimizing the sum of the elements below the matrix diagonal, which would give rise to the "min" formulation of the LOP. Throughout the remainder of this paper, we deal with the LOP formulation that maximizes the cost of the direct edges unless otherwise stated.

The LOP can be formulated by the following linear integer programming problem:

$$\text{Maximize} \sum_{i \in V} \sum_{\substack{j \in V \\ j \neq i}} c_{ij} x_{ij} \quad (2)$$

$$\text{subject to} \; x_{ij} + x_{ji} = 1 \qquad \forall\, i, j \in V, i \neq j \quad (3)$$

$$x_{ij} + x_{jk} + x_{ki} \leqslant 2 \quad \forall\, i, j, k \in V, i \neq j \neq k \neq i \quad (4)$$

$$x_{ij} \in \{0, 1\} \qquad \forall\, i, j \in V. \quad (5)$$

* Corresponding author.
E-mail addresses: sakuraba@usp.br (C.S. Sakuraba), dronconi@usp.br (D.P. Ronconi), egbirgin@ime.usp.br (E.G. Birgin), yagiura@nagoya-u.jp (M. Yagiura).

http://dx.doi.org/10.1016/j.eswa.2015.01.053
0957-4174/

In the model (2)–(5), $x_{ij} = 1$ if edge (i,j) is a direct one and $x_{ij} = 0$ otherwise. The objective function (2) is equivalent to the sum of the cost of the direct edges (1). Constraints (3) and (4) assure that the set of direct edges does not have 2 and 3-dicycles, which also prevents k-dicycles for $k \geqslant 4$ (Grötschel, Jünger, & Reinelt, 1984; Grötschel, Jünger, & Reinelt, 1985).

Besides the economic context, the LOP has various real world applications in many different fields, such as the aggregation of individual preferences in marketing, and the most probable chronological ordering of potteries in archeology (Grötschel et al., 1984), among others. In the sports domain, instances of the LOP corresponding to point differential matrices of 347 teams in NCAA college basketball are described in Sukegawara, Yamamoto, and Zhang (2011), while data corresponding to the results of ATP tennis tournaments with up to 452 players are available at <http://www.optsicom.es/lolib/#instances>. The increasing amount of information available nowadays and the need for solutions that integrate more levels of the production chain demand solutions for problems of a different scale. In fact, very large-scale instances of several classic problems, as for example instances of the set covering problem with more than a million columns and several thousand rows, are available at the OR-Library. Analysing instances of the LOP containing information of hundreds of sectors or classifying data from a pool with thousands of alternatives are far from being unrealistic situations. Therefore there is an imminent need for algorithms capable of dealing with large-scale instances.

Known to be an NP-hard problem (Karp, 1972; Garey & Johnson, 1979) (the proof of NP-completeness was given for the feedback arc set problem, which is equivalent to the LOP), the LOP has been extensively studied in the literature and many exact and heuristic methods have been proposed to solve it. Some results, conjectures and open problems dealing with the combinatorial and algorithmic aspects of the LOP are analyzed in Charon and Hudry (2007). It is possible to solve instances of the LOP with up to seventy-five vertices using its MIP formulation and the commercial solver CPLEX. However, instances with hundreds of vertices or more were handled in the literature so far by using heuristic and metaheuristic approaches. Next we provide a brief description of some relevant elements of these methods.

Chanas and Kobylanski (1996) suggested a multi-start algorithm (CK) that utilizes the symmetric property of LOP, i.e. if the permutation $\pi = (\pi(1), \pi(2), \ldots, \pi(n))$ is an optimal solution to the maximization problem (2)–(5) then $(\pi(n), \pi(n-1), \ldots, \pi(1))$ minimizes the objective function (2). The authors applied a local search based on an insertion mechanism and, once a local optimal solution is found, the process is re-started from its reverse permutation. In Laguna, Marti, and Campos (1999) the insertion move is also considered within four different variants of the tabu search method: (i) a basic procedure that alternates between an intensification and a diversification phase (TS); (ii) TS associated with path relinking; (iii) TS and a long-term diversification, and (iv) TS with path relinking and a long-term diversification. Computational tests suggested that version (iv) outperforms the CK algorithm. It should be pointed out that the long-term diversification was inspired by the reverse operation developed by Chanas and Kobylanski (1996). A scatter search algorithm was proposed in Campos, Glover, Laguna, and Martí (2001). It uses a frequency-based memory to guide the construction of a diversified initial reference set improved by the best neighborhood search developed in Laguna et al. (1999). Furthermore, a specific solutions combination method was developed for the LOP problem through the use of a min–max rule. A study about variants of the variable neighborhood search (VNS) to solve the LOP problem was carried in Garcia, Pérez-Brito, Campos, and Martí (2006). A hybrid version of VNS, where the local search is replaced with the short-term tabu search proposed in Laguna et al. (1999), is also analyzed in Garcia et al. (2006). In Huang and Lim (2003) and Schiavinotto and Stützle (2004) genetic algorithms are coupled with local searches. For the generation of new solutions Huang and Lim (2003) applied classical crossover and mutation operators associated with a local search that uses the insertion move and a first-fit search strategy. In the memetic algorithm (MA) proposed by Schiavinotto and Stützle (2004), where some additional crossover operators were evaluated, numerical results pointed out that the classical operators (cycle and order-based) performed best. Schiavinotto and Stützle (2004) also presented an iterated local-search (ILS) algorithm whose main components are: the perturbation scheme (based on interchange moves), the local search (ls_f), and the acceptance criterion (automatic tuning process). A common component of both proposed methods is the use of the ls_f procedure that uses the insertion move with the best search strategy and is able to evaluate the insert neighborhood in $O(n^2)$.

More recently, Kröemer, Platos, and Snasel (2013) proposed a bio-inspired metaheuristic based on artificial immune systems for the linear ordering problem. The proposed method consists in a modified B-cell algorithm based on clonal selection and it is quite similar to a GA method. A population of candidate solutions evolves by cloning, hypermutation, and selection. Numerical experiments, considering a library of small instances with known optimal solutions named LOLIB and comparing the introduced method with a GA and a differential evolution (DE) metaheuristic, are presented. In Ceberio, Hernando, Mendiburu, and Lozano (2013), the authors investigate instances' features that can provide useful insights into the difficulties of tackling the problem by applying local procedures associated with the insert neighbourhood. Ye, Wang, Lü, and Hao (2014) presents a multi-parent memetic algorithm that integrates a multi-parent recombination operator for generating offspring solutions and a distance-and-quality based criterion for the pool updating.

A comprehensive survey about existing metaheuristic approaches for the LOP problem was presented in Martí, Reinelt, and Duarte (2012), where computational experiments are reported. A comparison among the existing methodologies indicates that MA is the best method followed by ILS. The authors suggest that this good behavior may be due to the efficient implementation of the MA local search that, by calculating each neighbour solution in constant time, reduces the total cost of evaluating the insert neighborhood to $O(n^2)$. Moreover, the authors highlight that this local-search move complexity issue makes a significant difference in the overall performance of local-search methods and on metaheuristics that use the local-search strategies in their composition.

In this paper, we develop metaheuristic algorithms based on the TREE local-search strategy recently introduced in Sakuraba and Yagiura (2010), whose time complexity for considering the whole neighbourhood of a given solution is $O(n + \Delta \log \Delta)$, where Δ is the maximum degree of the graph. Computational experiments presented in Sakuraba and Yagiura (2010) show that the TREE local search outperforms the ls_f local-search strategy introduced in

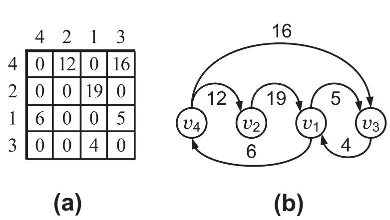

Fig. 1. (a) Matrix and (b) graph representations of the same solution to an instance of the LOP.