European Journal of Operational Research 241 (2015) 686–696

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Discrete Optimization

The linear ordering problem revisited

Josu Ceberio[a,*], Alexander Mendiburu[b], Jose A. Lozano[a]

[a] Intelligent Systems Group, Department of Computer Science and Artificial Intelligence, University of the Basque Country UPV/EHU, 200018 Donostia, Spain
[b] Intelligent Systems Group, Department of Computer Architecture and Technology, University of the Basque Country UPV/EHU, 200018 Donostia, Spain

ARTICLE INFO

Article history:
Received 19 December 2013
Accepted 21 September 2014
Available online 2 October 2014

Keywords:
Combinatorial optimisation
Linear ordering problem
Local optima
Insert neighbourhood

ABSTRACT

The Linear Ordering Problem is a popular combinatorial optimisation problem which has been extensively addressed in the literature. However, in spite of its popularity, little is known about the characteristics of this problem. This paper studies a procedure to extract static information from an instance of the problem, and proposes a method to incorporate the obtained knowledge in order to improve the performance of local search-based algorithms. The procedure introduced identifies the positions where the indexes cannot generate local optima for the insert neighbourhood, and thus global optima solutions. This information is then used to propose a restricted insert neighbourhood that discards the insert operations which move indexes to positions where optimal solutions are not generated.

In order to measure the efficiency of the proposed *restricted insert neighbourhood* system, two state-of-the-art algorithms for the LOP that include local search procedures have been modified. Conducted experiments confirm that the restricted versions of the algorithms outperform the classical designs systematically when a maximum number of function evaluations is considered as the stopping criterion. The statistical test included in the experimentation reports significant differences in all the cases, which validates the efficiency of our proposal. Moreover, additional experiments comparing the execution times reveal that the restricted approaches are faster than their counterparts for most of the instances.

1. Introduction

The Linear Ordering Problem (LOP) is a classical combinatorial optimisation problem which has received the attention of the research community since it was studied for the first time by Chenery and Watanabe (1958). Garey and Johnson (1979) demonstrated that the LOP is an *NP-hard* problem, thereby evidencing the difficulty of solving the LOP instances up to the optimality. However, due to its numerous applications in diverse fields such as archeology (Glover, Klastorin, & Klingman, 1972), economics (Leontief, 2008), graph theory (Charon & Hudry, 2007), machine translation (Tromble & Eisner, 2009) or mathematical psychology (Kemeny, 1959), we can find a wide variety of papers that have dealt with the LOP by means of exact, heuristic and metaheuristic strategies.

Among the exact methods, the most meaningful include Branch and Bound (Kaas, 1981; Charon & Hudry, 2006), Branch and Cut (Grötschel, Jünger, & Reinelt, 1984) and Cutting Plane algorithms (Mitchell & Borchers, 1996, 2000). These methods, as Schiavinotto and Stützle (2004) highlighted, behave competitively for instances from specific benchmarks with up to a few hundred columns and rows. However their computation time increases strongly with the size of the instances, and thus, it is not possible to solve large instances in a reasonable time span. Beyond the exact proposals, pioneering works proposed constructive heuristics (Chenery & Watanabe, 1958; Aujac, 1960; Becker, 1967). Such approaches were later outperformed by the advances produced in metaheuristic optimisation. Proof of this are the solutions based on Local Search (Kernighan & Lin, 1970; Chanas & Kobylanski, 1996), Genetic Algorithms (Charon & Hudry, 1998), Genetic Programming (Pop & Matei, 2012), Tabu Search (Laguna, Marti, & Campos, 1999), Scatter Search (Campos, Glover, Laguna, & Martí, 2001), Variable Neighborhood Search (Garcia, Pérez-Brito, Campos, & Martí, 2006), Ant Colony Optimisation (Chira, Pintea, Crisan, & Dumitrescu, 2009; Pintea, Crisan, Chira, & Dumitrescu, 2009), and recently Estimation of Distribution Algorithms (Ceberio, Mendiburu, & Lozano, 2013).

According to a recent review of Martí, Reinelt, and Duarte (2012), the Memetic Algorithm (MA) and the Iterated Local Search (ILS) proposed by Schiavinotto and Stützle (2004), are the algorithms that currently shape the state-of-the-art of the LOP. The MA is a hybrid algorithm which combines the canonical structure of a Genetic Algorithm with a high presence of local search procedures, either in the initialisation of the population or in the evolutionary process itself. On the other hand, the ILS is a strategy that iteratively applies

* Corresponding author. Tel.: +34 943 018 070.
E-mail addresses: josu.ceberio@ehu.es, jceberio001@gmail.com (J. Ceberio), alexander.mendiburu@ehu.es (A. Mendiburu), ja.lozano@ehu.es (J. A. Lozano).

http://dx.doi.org/10.1016/j.ejor.2014.09.041
0377-2217/

a local search algorithm to a single solution. When the process gets trapped in a local optimum solution, the ILS applies a perturbation to the current solution, and continues with the optimisation process until a termination criterion is satisfied. Both algorithms include an efficient implementation of a greedy local search algorithm with the insert neighbourhood designed specifically to solve the LOP.

As seen for most of the combinatorial optimisation problems, the hardness of solving a specific instance is not only limited to the size of this, but also to other additional parameters, unknown in most cases. In this regard, the community has tried to better understand the characteristics of the LOP that determine the difficulty of the instances, and, similarly, has tried to identify the features that could be useful to guide the algorithms throughout the optimisation process. In this sense, Schiavinotto and Stützle (2004) sketched out the properties that could somehow characterise the hardness of the LOP instances. The authors defined the *sparsity*, *variation coefficient (VC)*, *skewness* and *fitness distance correlation* as measures of instance hardness, and showed that real-life instances, which are apparently more difficult than artificial ones, present significant differences in *sparsity*, *VC* and *skewness* with respect to random benchmark instances. Nevertheless, the relation between the mentioned properties, and the suitability of the MA and the ILS to solve the LOP is not straightforward.

In the same research line, Betzler, Guo, Komusiewicz, and Niedermeier (2011) published a detailed work on the parameterised complexity for intractable median problems, and particularly on the Kemeny ranking problem, which can be seen as a subclass of LOP. Although the parameterised complexity studied in the cited work is of great relevance, the analysis of Betzler et al. (2011) stands on a specific property of the Kemeny, which does not hold for the general LOP, and thus, the extension of the parameterised complexity to the LOP is not straightforward.

The aforementioned works and the absence of a detailed work that performs an in-depth analysis of the LOP motivated this paper. In this work we study the properties that the optimal solutions of the LOP hold in the framework of local search algorithms, placing special emphasis on the position where the indexes are placed, and identifying the role of the associated matrix entries of the instance in the generation of local optima.

The paper is divided into two parts: first, we provide a detailed description of the LOP, introducing definitions and theorems that study the structure of the problem with respect to the optimality of the solutions in the context of local search algorithms. Particularly, we emphasise the influence that the positions of the indexes that compound a solution have when generating local optima solutions. As a result of the theoretical study, a *restricted* version of the insert neighbourhood is proposed. This neighbourhood discards specific insert operations that involve moving indexes to positions at which they cannot generate local optima solutions. The theoretical analysis demonstrates that these insert operations will never be the operations that most improve the solution in the neighbourhood.

The second part of the paper is devoted to demonstrating the validity of the *restricted insert neighbourhood*. In this sense, we develop a *restricted* version of the two best performing algorithms for the LOP: the MA and the ILS. Experimental results show that the restricted versions of the algorithms outperform the classical designs in 90 percent and 93.3 percent of the executions respectively, obtaining the same results for the rest of the cases. Moreover, additional experiments devoted to measure the execution time needed to perform a given number of iterations show that the *restricted* approaches are faster than the classical versions for most of the instances.

The remainder of the paper is organised as follows: in the next section, the definition of the LOP is described. In Section 3, the structural analysis of the LOP is introduced placing special emphasis on the contribution of the indexes to the objective function. Next, in Section 4 the optimality of the LOP solutions is described in the context of local search algorithms, and in particular for the insert neighbourhood system. Section 5 is devoted to investigating the basis for the restricted insert neighbourhood system. In order to demonstrate the validity of the introduced analysis, a complete experimental study is introduced in Section 6. Finally, some conclusions and ideas for future work are drawn in Section 7.

2. The linear ordering problem

Given a matrix $B = [b_{kl}]_{n\times n}$ of numerical entries, the linear ordering problem consists of finding a simultaneous permutation σ of the rows and columns of B, such that the sum of the entries above the main diagonal is maximized (or equivalently, the sum of the entries below the main diagonal is minimized). The equation below formalizes the LOP function:

$$f(\sigma) = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} b_{\sigma_i\sigma_j} \quad (1)$$

where σ_i denotes the index of the row (and column) ranked at position i in the solution σ.[1] This representation of the LOP is also known as the *triangulation problem of input-output matrices*. Although alternative representations of the problem can be found in Martí and Reinelt (2011) and Charon and Hudry (2007), due to the theoretical simplicity and readability of the exposed approach, in the remainder of the paper the *triangulation* representation will be considered.

Example 2.1. Let us introduce an example for a $n = 5$ LOP instance which will be used throughout the paper.[2] In Fig. 1, three different solutions, e, σ and σ^* are described. The initial matrix is represented by the identity permutation $e = (1, 2, 3, 4, 5)$ (see Fig. 1a), and its fitness, $f(e)$, is 138. The solution $\sigma = (2, 3, 1, 4, 5)$ introduces a different ordering of the indexes that provides a solution better than e (see Fig. 1b), $f(\sigma)$ is 158. The optimal solution for this example is given by $\sigma^* = (5, 3, 4, 2, 1)$ (see Fig. 1c), with fitness $f(\sigma^*) = 247$.

3. Analysis of the problem

In this section, we analyse the LOP by explaining the association between the indexes in σ and the arrangement of the b_{kl} entries of the matrix B. In addition, we describe the fitness variation that provokes changing the position of an index within σ, and the role of the b_{kl} entries in this regard. As necessary background to understand the latter content of the paper, in the following list we outline some meaningful properties of the LOP that define the association between the indexes in σ, and the b_{kl} entries in the B matrix.

For any permutation of indexes σ of size n and a matrix B of size $n \times n$:

- Every index $\sigma_i = k$, $i = 1, \ldots, n$, has associated $2(n-1)$ entries of B: $n-1$ from row k and $n-1$ from column k.
- The set of associated entries of every index $\sigma_i = k$, $i = 1, \ldots, n$, can be organised in pairs, i.e. every entry in row k, $b_{k\sigma_j}$ (where $j = 1, \ldots, n$), has a pair in column k, $b_{\sigma_j k}$, symmetrically located with respect to the main diagonal.
- All the pairs of entries associated to index $\sigma_i = k$, $\{b_{k1}, b_{1k}\}, \ldots, \{b_{kn}, b_{nk}\}$, remain associated to this index regardless of its position and the position of the rest of the $n-1$ indexes.
- Every entry $b_{\sigma_i\sigma_j}$ is associated to two indexes, σ_i and σ_j.
- For every pair $\{b_{\sigma_i\sigma_j}, b_{\sigma_j\sigma_i}\}$ of entries, one entry is always located above the main diagonal, and the other entry is located below, thereby bounding the best fitness contribution of this pair to

[1] From now on, σ will denote any permutation in $\mathbb{S}_n$, and e will stand for the identity permutation $(1, 2, \ldots, n)$ of size n. In addition k and l will denote the indexes within a permutation σ, and i, j and z will be used to identify the positions of σ.

[2] This example was extracted from Martí and Reinelt (2011).