Solving Project Scheduling Problems using Estimation of Distribution Algorithm with Local Simplex Search

1Yan-Jun Shi, 1Guo-Jiang Shen and 3Wang Chen
1School of Mechanical Engineering, Dalian University of Technology, Dalian 116024,
People’s Republic of China
2China North Vehicle Research Institute, Beijing 100072, People’s Republic of China

Abstract: The Resource-Constrained Project Scheduling Problem (RCPSP) is an NP-hard problem in information engineering. The activities of a project have to be scheduled for satisfying all the precedence and resource constraints. We presented a heuristic algorithm (EDAS) to deal with this problem which employed an estimation of distribution algorithm (known as EDA) and improved the local search capacity with a simplex search. In this algorithm, the EDA firstly searched the solution space and generated activity lists to provide the initial population; then, the EDA selected the sample solutions to build a probability distribution model. The new individual was generated by sampling this model. The simplex search was used to enhance the local search capacity of the EDA. Compared with state-of-the-art algorithms available in the literature, we showed the effectiveness of this approach empirically on the standard benchmark problems of size J60 and J120 from PSPLIB.

Key words: Resource-constrained, project scheduling, estimation of distribution algorithm, simplex method

INTRODUCTION

The Resource-Constrained Project Scheduling Problem (RCPSP) (Hartmann and Briskorn 2010) is a classical NP-hard problem in scheduling. The problem is widely applicable in project management, construction engineering, software development and production scheduling. The objective is to schedule the activities of a project so as to minimize the project makespan, subject to precedence and resource constraints. This problem has been well researched for over four decades. Previous studies have proposed several meta-heuristics which have been successfully applied to solve large scale problems with better robustness and higher calculating speed. Chen (2010) developed a mixed binary integer programming model to optimally solve this kind of NP-hard problems. Xie (2009) proposed a serial scheduling and parallel scheduling algorithm which firstly formulated the problem to a nonlinearity programming problem and then transformed it into an integer programming problem based on the piece linear rate-distortion model. Ziarati et al. (2011) investigated the application of bee algorithms RCPSP. In the study of Chen (2011), the justification technique was combined with PSO as the proposed Justification Particle Swarm Optimization (JPSO) in solving RCPSP. Zhu et al. (2010) presented an effective Forward and Reverse Schedule Generation (FRSG) method with a novel complete local search to find the solution with the best objective value. Benedict and Vasudevan (2008) proposed an Improved Tabu Search Algorithm for the project scheduling problems. In our previous studies, we proposed an efficient algorithm ACOS, combined a local search strategy, Ant Colony Optimization (ACO) and a Scatter Search (SS) in an iterative process for solving RCPSP (Chen et al., 2010).

Different from the previous studies, we aim at showing the feasibility and effectiveness of the estimation of distribution algorithm for the RCPSP. Estimation of Distribution Algorithms (EDA) (Larrañaga and Lozano, 2002) are evolutionary algorithms based on estimation of and sampling from probabilistic models, have been successfully applied to the combinatorial optimization problems (Campelo, Guimarães et al., 2009; Eddaly et al., 2009; Jarboui et al., 2009; Naeem and Lee, 2009; Luo et al., 2010; Meiguins et al., 2010; Rodrigues and Yamashita, 2010; Wang et al., 2010a; Xu et al., 2010; Zhong et al., 2010; Zhang and Li, 2011). Campelo et al. (2009) introduced an approach for the design of electromagnetic devices based on the use of estimation of distribution algorithms coupled with approximation-based local search around the most promising solutions. Eddaly, Jarboui et al. (2009) propose

Corresponding Author: Yan-Jun Shi, School of Mechanical Engineering, Dalian University of Technology, Dalian 116024, People’s Republic of China Tel:+86 0411 84709130 Fax:+86 0411 84708412

1374
an Estimation of Distribution Algorithm for solving a flow-shop scheduling problem with buffer constraints. Rodrigues and Yamashita (2010) used the EDA algorithm for minimizing resource availability costs in project scheduling. In the traditional evolutionary computation, new candidate solutions are often generated by combining and modifying existing solutions in a stochastic way. The underlying probability distribution of new solutions over the space of possible solutions is usually not explicitly specified. In the EDA, the representation of the population is replaced with a probability distribution over the choices available at each position in the vector that represents a population member. So a population may be approximated with a probability distribution and new candidate solutions can be obtained by sampling this distribution. This algorithm may have several advantages, such as, avoiding premature convergence and being a more compact representation.

But, as a global search algorithm, EDA does not perform well for the large-scale problems with a slow convergence speed. It’s known that an effective algorithm not only can explore the solution space sufficiently and also can find the elites by local search. After the analysis of the RCPSP problems, this study improves the local search capacity of the EDAS algorithm with a simplex search. The simplex method (Klee and Minty et al., 1970) is a popular algorithm for numerically solving linear programming problems. It can be viewed as a family of combinatorial local search algorithms on the boundary of a solution space (Cosio et al., 2010; Li et al., 2010; Osgood et al., 2010; Song et al., 2010; Wu et al., 2010; Xu et al., 2010). The search moves from one point of the solution space to a neighboring one with a better objective value, according to a chosen rule. Today, the simplex method is still widely used and remains important in practice, particularly in combinatorial optimization and local search.

In this study, we will describe how to integrate the EDA and simplex approach. Our empirical test showed that our hybrid approach performed well on the standard benchmark problems of size J60 and J120 from PSPLIB.

THE RCPSP MODEL

The Resource Constrained Project Scheduling Problem (RCPSP) is stated as follows. For the project A, n+2 activities I = 0,1,2,...,n,n+1 and r renewable resources are given. A constant amount of R_k units of resource k is available in each time unit. The duration of an activity i is given by d_i. During this time period, a constant amount of r_k units of resource k is occupied by the activity i. The dummy activities 0 and n+1 represent the beginning and the end of the project, where d_0 = d_{n+1} = 0 and r_k = r_{n+1}. Parameter d, r_k and R_k are assumed to be non-negative integer values. The activities are interrelated by two types of constraints: (1) precedence constraints prevent activity i from being started before all its predecessor activities P_i have been completed and (2) the sum of the resource requirements for resource k in anytime period cannot exceed R_k. The optimized objective of the RCPSP is to find precedence and resource feasible completion times for all activities such that the make-span of the project is minimized.

Let the completion time of activity i be denoted as D_i, then the completion times of schedule S are denoted as D_1, D_2,..., D_n. The conceptual model of the RCPSP can be described as follows:

\[ \min D_{n+1} \]  \hfill (1)

\[ \text{st. } D_i \leq F_i - d_i, \quad i = 1, 2, ..., n+1; \]  \quad \forall h \in P_i \hfill (2)

\[ \sum_{k \in K(t)} r_k \leq R_k \quad \forall k \in K, t \geq 0 \]  \hfill (3)

\[ D_i \geq 0 \quad \forall i = 1, 2, ..., n+1 \]  \hfill (4)

The objective function given by Eq. 1 minimizes the make-span of the schedule. Equation 2 enforces the precedence constraints between activities, while Eq. 3 enforces the resource limitation constraint, where A(t) = i |d_i - s_i = D_i, i.e., the set of activities being processed (active) at time t. Finally, Eq. 4 describes the constraint of the decision variables.

THE EDAS ALGORITHM

The main flowchart of EDAS algorithm: In this study, we integrate the estimation of distribution algorithms and the simplex method to construct the Estimation of Distribution Algorithms with Simplex method (EDAS) for solving the RCPSP problems. Figure 1 shows the main flowchart of the EDAS. In the initial stage, the parameters and population is initialized. The tournament selection is employed to obtain the selected solutions. Then the EDA explicitly extracts global statistical information from the selected solutions by (often called parents) and build a posterior probability distribution model of promising solutions, based on Gaussian model with diagonal covariance matrix (GM/DCM) (Larranaga and Lozano, 2002; Wang et al., 2010a,b). New solutions are sampled from the model thus built and fully or in part replace
the old population. In addition, the simplex method is used to search local optima in each iteration and the limitation of EDA such as poor local search ability and premature is largely enhanced. The SM searches the local neighborhood of the current optimal feasible solution to improve the velocity of convergence.

**The representation and decoding procedure for the EDAS algorithm:** The representation and decoding procedure for RCPSP plays an important role in algorithm performance. We herein employ the random key representation (Shi et al., 2010), a solution was represented by a random sequence and denote $s_i$ as the number of activities in the sub-problem $S$:

$$X_i = (x_{i1}, x_{i2}, \ldots, x_{in}), (s = 1, 2, \ldots, S), 0 < x_{ij} < 1 (j = 1, 2, \ldots, n_i)$$

In random key representation, schedule generation schemes (SGS) are the core decoding procedures for the RCPSP (Kolisch and Hartmann, 1999). Two different SGSs are available in the literature: (1) the serial SGS that performs activity-incrementation and (2) the parallel SGS that performs time-incrementation. Considering the advantages and disadvantages of both schemes, we use
a random real number, \( r \in (0, r_{\text{max}} < 1) \) to select the SGS with which to construct the schedule, as defined by:

\[
SGS = \begin{cases} 
\text{serial-SGS, if } r < 0.5 \\
\text{parallel-SGS, if } r > 0.5 
\end{cases} \tag{5}
\]

The sample solution selection method for the EDAS algorithm: It is important to select the sample solution for applying the EDA. We employ the tournament selection as a countermeasure. The tournament selection randomly selects two individuals in a population and the individual with better fitness will be the sample solution. Repeat the process until the number of the sample solutions meets the requirements. In this paper, the number of the sample solutions is only half of that of the population. We denote NP as the number of the population, so the number of the sample solutions is NP/2.

The probability distribution model for the EDAS algorithm: This step is to build a probability distribution model. Since the distribution of categorical variables is Gauss distribution or similar Gauss distribution in the realistic situation, so we employ Gaussian model with diagonal covariance matrix (GM/DCM) to build a probability distribution model. GM/DCM considers the correlation between random variables and uses every element of the matrix to represent the covariance among different components of a random vector. This method describes the correlations among the different design variables more reasonably and these correlations are applied to guide the further search.

Each design variable obeys the normal distribution, so the joint probability density function of the sample solution (n dimensions random vector \( \mathbf{x} = (x_1, x_2, ..., x_n) \) is:

\[
f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp \left( -\frac{1}{2}(\mathbf{x} - \mu)^T \mathbf{C}^{-1}(\mathbf{x} - \mu) \right) \tag{6}
\]

We denote \( \mathbf{C} \) as covariance matrix of \( n \) dimensions random vector, \(| |C| | \) as the determinant of covariance matrix and \( \mathbf{C}^{-1} \) as the inverse matrices of \( \mathbf{C} \). We can estimate the mean \( \mu \) and covariance matrix \( \mathbf{C} \) by Maximum Likelihood Estimate (MLE) (Laurence and Chromy, 2010, Yin et al., 2011):

\[
\mu = \frac{1}{(NP/2)} \sum_{i=1}^{NP/2} x \tag{7}
\]

\[
\mathbf{C} = \frac{1}{(NP/2)} \sum_{i=1}^{NP/2} (x_i - \mu)(x_i - \mu)^T \tag{8}
\]

The mean \( \mu \) and covariance matrix \( \mathbf{C} \) need to be re-calculated in every iteration step.

The sampling method for the EDAS algorithm: For generating the new individual, the EDA here is implemented for sampling the joint probability density function which is employed for describing the probability distribution model. As for each of the random variables, sampling method is used to generate random numbers. The vector formed by these random numbers, will be a new individual. Repeat this procedure until the whole new population is completed. In the GM/DCM, each dimension of the design variables obeys the normal distribution, \( x_i \sim \text{N}(\mu, \sigma) \). This paper employs Box-Muller formula as the sampling method:

\[
r_{\text{max}} = \sqrt{-2 \ln r_1 \sin (2\pi r_2)} \tag{9}
\]

\[
x_i = r_{\text{max}} \cdot \sigma_i + \mu_i \tag{10}
\]

In Eq. 9, \( r_1 \) and \( r_2 \) are random numbers between 0 and 1 while random number \( r_{\text{max}} \) obeys the normal distribution, \( \text{N}(0,1) \). And based on Eq. 10, the sample point \( x_i \) is determined which obeys the normal distribution, \( x_i \sim \text{N}(\mu, \sigma) \). By this way, the complete solution is obtained.

The local search strategy using simplex search: The simplex search is very suitable for nonlinear function optimization without any gradient information and coding operation. The process is composed of three basic steps are reflection, compression and expansion. The iterative processes are follows:

Step 1: Initialize the compressibility \( \lambda \), the expansion coefficient \( \beta \) and the tolerance \( \varepsilon \).

Step 2: Generate randomly \( n+1 \) points, \( x_0, x_1, ..., x_n \) to form a simplex.

Step 3: Calculate the values of function \( f(x_i) \), \( i \in 0, 1, ..., n \) to get the best point \( x_0 \), the worst point \( x_i \). We denote the fitness of them as follow:

\[
f_i = f(x_i) = \max \left\{ f(x_0), ..., f(x_i) \right\}
\]

\[
f_i = f(x_i) = \min \left\{ f(x_0), ..., f(x_i) \right\}
\]

Step 4: Calculate the centre point \( x_c \) of \( x_0, ..., x_{n+1} \) except the point \( x_0 \)

\[
x_c = \frac{\sum_{i=0}^{n+1} x_i}{n+1}
\]

and the reflection point \( x_r, x_r = 2x_c - x_0 \).

Step 5: If \( f(x_r) \geq f_0 \), compress the pre-processed data, \( x_i \leftarrow x_i + \lambda(x_c - x_0) \). Then calculate \( f_c = f(x_c) \) and go to Step 7; if \( f_c < f_0 \) go to Step 6.
Step 6: Carry out the expansion, $x_i = x_i + \mu(x_i - x_i)$ and calculate $f_i = f(x_i)$. If $f_i \leq f_k$ implement the following assignments $x_i = x_i$ and $f_i = f_i$; else $x_i = x_i$ and $f_i = f_k$.

Step 7: If $f_i < f_k$, substitute $x_i$ for $f_i$ and that $f_i$ for $f_k$. Put the new point $x_i$ with the other $n$ points to form a new simplex. Retrieve the best point $x_i$ and the worst point $x_w$, then go to Step 4; if $f_i \geq f_k$, go to Step 8.

Step 8: If $f_i - f_k < c \cdot |f_k|$, end the process and set $x^* = x_i$ and $f^* = f_i$; else shorten the step size and follow the formula $x_i = (x_i + x_w)/2$, $i = 0, 1, ..., n$, then go to Step 3, continue the calculation.

THE EXPERIMENTS

We applied the EDAS to the standard instance sets J60 (480 instances each with 60 activities) and J120 (600 instances each with 120 activities) (Kolisch and Sprecher, 1997). These sets are available in the well known PSPLIB along with the optimum, or best known values that have been obtained by various authors over the years. In J60 and J120 the optimal make-spans for some instances are not known and only upper bounds (current best solutions) and lower bounds are provided. For a fair machine-independent comparison, we limited the number of generated and evaluated schedules in EDAS to 1000, 5000 and 50,000, respectively. EDAS was implemented in JAVA 1.6 and all the experiments were performed on a PC with 1.86 GHz CPU and 2 GB RAM running the Windows XP operating system.

In the preliminary experiments, we employed a trial-and-error strategy to obtain suitable control parameter values for each algorithm. We settled on the following control parameters: the number of the population, $N_P = 100$ and the number of sub-problems, $S = 4$.

Table 1 and 2 gave the performances of EDAS and several RCPSP heuristics from the literature with 1000, 5000 and 50,000 evaluated schedules. This study compared EDAS with GAPS (Mendes et al., 2009), GA-DBH (Debels and Vanhoucke, 2007), GA-hybrid-FBI (Valls et al., 2003) and GA-FBI (Valls et al., 2005).

Table 1 listed the lower bound results for instances with 60 activities. The EDAS ranked 4th for 1000 schedules and both 2nd for 5000 and 50,000 schedules, respectively. In this situation, the performance of the GA-FBI method is the worst of the five kinds of algorithms. The GA-DBH showed better results than the EDAS which meant that the GA-DBH is very suitable for project scheduling with 60 activities.

Table 2 listed the results for J120, where the EDAS algorithm ranked 3rd for 1000 schedules. As for 5000 schedules, the EDAS ranked 1st which was equal to the GA-DBH algorithm. The EDAS ranked 1st for 50,000 schedules showed that EDAS can produce better solutions for the large-scale problems J60 and J120. That is to say, the EDAS algorithm can solve the large-scale RCPSP problems effectively.

The reason for the improved performance is that EDAS combines the advantages of the GM/DCM and the simplex local search method. The GM/DCM describes the correlations among the different design variables more reasonably and the correlations are applied to guide the further search. The simplex method is used to search local optima in each iteration to fill the gaps of the EDAS such as poor local search ability and the lower velocity of convergence.

CONCLUSIONS

In this study, we proposed a heuristic algorithm combining the Estimation of Distribution Algorithm (EDA) and the simplex local search strategy to solve resource-constrained project scheduling. The experimental results of the EDAS algorithm were compared with the state-of-the-art heuristics on the standard instance sets and from the PSPLIB. The experimental results indicated that, EDAS performed well compared with other heuristic algorithms for the large-scale RCPSP problems. This study also showed that the EDAS is a promising method for solving the RCPSP. Further studies will focus on enhancing the applicability and efficiency of the proposed EDAS algorithm for various complex real-world scheduling problems.
ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of P.R. China (Grant No. 50975033, 50975039).

REFERENCES


