Constraint-based techniques for Master thesis defense scheduling problem with multiple objectives

ABSTRACT

General Terms

Applications

Keywords

Constraint Programming, Local Search, Optimization

1. INTRODUCTION

We consider in this paper a real-world combinatorial application emerging in the education management in Vietnam: the problem of scheduling of the master defenses at universities. A set of Master students at a given school having sufficient credits will be organized into a defense of their thesis session. Each student is associated with a supervisor, and the supervisor will introduce two reviewers for the thesis of this student and these reviewers will be strongly encouraged to attend his defense session. The jury of each defense requires three other professors including a president, a secretary and an examiner. The policy of most of Vietnamese universities impose a condition that the supervisor of a student cannot attend his defense session, and the jury must consist three internal professors (i.e., professors that are inside the university) and two external professors (i.e., professors that are outside the university). There are also a set of rooms and time slots available for the defense. The organizer must specify, for each student, a jury, a room and a time slot for his defense satisfying a set of constraints described in the problem formulation section (see Section **??**).

1.1 Problem formulation

This section describes the formulation of the problem in a generic way. We are given

- a set of predefined subject domains $SD = \{1, \ldots, |SD|\}$
- a set of rooms $R = \{1, \ldots, |R|\}$ available for accommodating the defense sessions
- a set of time slots $SL = \{1, \ldots, |SL|\}$.
- a set of professors $P = \{1, \ldots, |P|\}$ including professors inside and outside the university, each professor $p \in P$ is associated with following information
  - a set of subject domains $D[p]$ he can follow ($D[p] \subseteq SD$). Each subject domain $d_j \in D[p]$ is associated with an evaluation function $f(p, d_j)$ which describes the suitability of the professor p and the subject domain d_j. In this paper, the value of $f(p, d_j)$ are pairwise different and are taken from $\{0, 1, \ldots, k-1\}, \forall i \in \{1, \ldots, k\}$. $f(p, d_j) = 0$ means that the professor p suits most with d_j.
  - a set of time slots he is available $T[p] \subseteq SL$
  - a boolean value $isHust[p]$ which is true if the professor p is inside the university
- a set of students $S = \{1, \ldots, |S|\}$ who will do the defense of their theses. Each student $s \in S$ is associated with following information
  - subject domain $d[s]$ of his thesis: $d[s] \in SD$
  - the supervisor of the thesis $Sup[s] \in P$
  - a set reviewers of the thesis $Rev[s] \subseteq P$

We must compute $|S|$ scheduling lines, each of which corresponds to a student. A scheduling line $sched[s]$ corresponding to the student s contains the following information

- the room $xr[s]$ and the time slot $xt[s]$ where the defense of the student s takes place
- a jury of the defense of the student s including J professors $xp[s, 1], \ldots, xp[s, J]$

The scheduling lines of $|S|$ students must satisfy the following constraints

- (C1): the members of a jury must be different: $xp[s, i] \neq xp[s, j]$, $\forall i \neq j \in \{1, \ldots, J\}, s \in S$
- (C2): the supervisor of a student cannot attend to his defense session: $Sup[s] \notin \{xp[s, 1], \ldots, xp[s, J]\}$
- (C3): at least k reviewers of a student must be present at his defense session: $|Rev[s] \cap \{xp[s, 1], \ldots, xp[s, J]\}| \geq k$
- (C4): a room cannot accommodate two defenses at a time slot: $xr[s_1] = xr[s_2] \Rightarrow xt[s_1] \neq xt[s_2]$

SoICT2011 , October 13-14, 2011, Hanoi, Vietnam.

- (C5): a professor cannot attend two defenses at a same time slot: $xt[s_1] = xt[s_2] \Rightarrow \{xp[s_1, 1], \dots, xp[s_1, J]\} \cap \{xp[s_2, 1]), \dots, xp[s_2, J]\} = \emptyset$

- (C6): the jury must contain J^{in} professors inside the university and J^{out} professors outside the university: $J^{in} + J^{out} = J$

$$\sum_{j=1}^{J} isHust[xp[s, j]] = J^{in}$$

- (C7): the time slot specified for each student's defense must be in the available time slots of professors attending his defense

$$xt[s] \in T[xp[s, j]], \forall s \in S, j \in \{1, \dots, J\}$$

- (C8): the subject domain of the thesis of a student must be in the set of subject domains of each professor attending his defense

$$d[s] \in D[xp[s, j]], \forall s \in S, i \in \{1, \dots, J\}$$

Three objectives are taken into account. The first objective is to maximize the suitability of the subject domains of professors with respect to the subject domain of the thesis they are attending. The second objective is to maximize the consecutiveness of the slots each professors attending. The third objective is to minimize the total number of change of the rooms of all the professors. We denote X the set of all decision variable of the formulation:

$$X = \cup_{s \in S} \{xr[s], xt[s], xp[s, 1], \dots, xp[s, J]\}$$

The first objective is modeled by the following function to be minimized:

$$F(X) = \sum_{s=1}^{S} \sum_{j=1}^{J} f(xp[s, j], d[s])$$

Formalizing the second and the third objectives in a direct way is not evident. For ease the formulation, we use auxiliary notation $rp(p, t)$ which represents the room where the professor p is assigned at time slot t. If the professor p is not assigned at time slot t, then $rp(p, t) = 0$ by convention. Clearly, $rp(p, t)$ can be derived from X ($\forall p \in P, t \in SL$). For each professors $p \in P$, we denote additional notations

- $ts(p) = \{t \in SL \mid rp(p, t) > 0\}$
- $l(p) = min(ts(p))$
- $u(p) = max(ts(p))$
- $C(p)$ denotes the consecutive evaluation of p

  $C(p) = |\{\langle t_1, t_2\rangle \mid t_1 < t_2 - 1 \wedge rp(p, t) = 0, \forall t_1 < t < t_2\}|$

- $S(p)$ denotes the shift evaluation of p

$$S(p) = \sum_{t=l(p)}^{u(p)-1} rp(p, t) \neq rp(p, t + 1)$$

Intuitively, $C(p)$ is the number of pairs of slots $\langle t_1, t_2\rangle$ ($t_1 < t_2 - 1$) where p is assigned to time slots t_1 and t_2 and is not assigned to any time slot t between t_1 and t_2.

Student	Room	Slot	R1	R2	Pres.	Secr.	Exa.
1	1	1	19	45	9	18	21
2	1	2	19	45	9	18	21
3	2	1	38	51	16	33	35
4	3	1	17	43	1	3	28
5	1	6	26	56	28	30	32
6	3	2	26	29	3	28	32

Table 1: Example of scheduling

Professor	rp						C	S
1	3	0	0	0	0	0	0	0
3	3	3	0	0	0	0	0	0
9	1	1	0	0	0	0	0	0
16	2	0	0	0	0	0	0	0
17	3	0	0	0	0	0	0	0
18	1	1	0	0	0	0	0	0
19	1	1	0	0	0	0	0	0
21	1	1	0	0	0	0	0	0
26	0	3	0	0	0	1	1	2
28	3	3	0	0	0	1	1	2
29	0	3	0	0	0	0	0	0
30	0	0	0	0	0	1	0	0
32	0	3	0	0	0	1	1	2
33	2	0	0	0	0	0	0	0
35	2	0	0	0	0	0	0	0
38	2	0	0	0	0	0	0	0
43	3	0	0	0	0	0	0	0
45	1	1	0	0	0	0	0	0
51	2	0	0	0	0	0	0	0
56	0	0	0	0	0	1	0	0

Table 2: Auxiliary information: rp, C, and S

rp, C, S are called derived information from X: When the value of X are specified, then rp, C, S are specified deterministically.

The second objective can now be modeled as the following function to be minimized:

$$C(X) = \sum_{p \in P} C(p)$$

The third objective is modeled by the following function to be minimized:

$$S(X) = \sum_{p \in P} S(p)$$

Example.

Table **??** gives an example of the defense scheduling with 6 students $\{1, 2, 3, 4, 5, 6\}$, 6 time slots $\{1, 2, 3, 4, 5, 6\}$, 3 rooms $\{1, 2, 3\}$ and 56 professors $\{1, \dots, 56\}$. Columns 2–3 present the room and the time slot of the defense. Columns 4–8 present 5 members of the jury.

Table **??** describes derived information from the scheduling solution in Table **??** in which columns 2–7, column 8, and column 9 respectively depict the rp, C, and S information of the professors (each row of the table corresponds to a professor). In this case, the values of the second and the third objective are 3 and 6.

The ideal solution would have zero values for all three objectives. $F(X) = 0$ means that for each student, the subject domain of his master thesis is the most suitable subject domain of all professors attending his defense. $C(X) = 0$ means that each professor attending the defense sessions in a sequence of consecutive time slots. Naturally, no professor desires to be interrupted before attending the subsequent defense. $S(X) = 0$ means that no professor has to shift the rooms during the time period he attends the defense sessions.

Clearly, this is a constraint optimization problem with multiple objectives. From realistic requirements, we handle three objectives in a lexical way: A solution X_1 is better than a solution X_2 if $(F(X_1), C(X_1), S(X_1)) < (F(X_2), C(X_2), S(X_2))$

1.2 Motivation and Objectives

As far as we know, this problem was not considered before. Traditionally, this real-world scheduling problem was realized by human, and it was always stressing and time-consuming. Moreover, from a modeling stand point, the problem is interesting and not trivial. From a computation stand point, finding optimal solution is really challenging. In this paper, we propose some algorithms based on Constraint Programming (CP) and Constraint-Based Local Search (CBLS) approaches for solving this problem. We experimentally compare them on real instances from the School of Information and Communication Technology (SoICT), Hanoi, Vietnam. We show that the CBLS algorithm is efficient in large instances where the CP algorithms cannot handle. Moreover, when the CP algorithm proof that no feasible solution exists, then the CBLS can provide a high quality solution and it can indicate which parts of the solution violate the constraint, and thus suggest the human how to correct these violating parts. For example, when the provided solution violates the time slots of some professors, then we can find and invite other professors that are not in the input list to attend the defense. The proposed constraint-based approach can easily be extended when other side constraints need to be added. Experimental results show the interest of the proposed algorithms.

1.3 Related work

Timetabling within a university context has long been recognized as difficult from both a theoretical and practical perspective. A huge variety of timetabling models have been described in the literature; they range from the weekly timetable of a school to the scheduling of courses, exams or some other events in a university [?]. These problems are viewed as educational timetabling problems which can be defined as the problem of assigning a number of events such as exams, courses to a given number of timeslots and rooms while satisfying a set of constraints [?, ?]. These constraints are usually classified into two types. Hard constraints must be satisfied in order to provide a feasible solution, whereas, soft constraints can be violated, but we try to satisfy them as far as possible. The quality of a timetable is measured based on how well the soft constraints have been satisfied.

Over the last twenty years, meta-heuristic approaches have been successfully applied to educational timetabling problems. For example, graph based heuristics [?, ?], tabu search [?], large neighborhood search [?], great deluge algorithms [?], hybrid algorithms [?], and population based algorithms including memetic algorithms [?], ant colony optimization [?], genetic algorithms [?], honey-bee mating optimization (HBMO) algorithms [?] have all been utilized. However these algorithms are often very specific and slight changes in the problem definition raise difficulties in the adaptation of the special purpose algorithms.

There is possibility of instantly changing already defined algorithms and making it adequate for real-world problems, but it is hard and it often lowers their effectiveness. One of the offspring of AI is Constraint Programming (CP), which offers flexibility by the formulation of constraints in a high-level language. Its main advantage is declaratively: a straightforward statement of the constraints serves as part of the program. This makes the program easy to modify, which is crucial in real-world problems. Constraint Programming has succeeded in solving standard benchmarks and real-world problems from the area of scheduling in general and the area of timetabling in particular [?, ?, ?]. The construction of timetabling problem differs greatly for different schools and educational institutions. Based on a specific problem model, the formulation presented contained characteristics of the timetabling problem found in many universities. Motivated by the above, this work proposes a constrained-based approach for a specific timetabling model on master thesis defense in universities.

Structure of the paper.

2. ALGORITHMS

We divide the algorithm into three steps: the first step $\mathcal{A}_1$ computes the solution optimizing the first objective function. In the second step $\mathcal{A}_2$, we compute the solution optimizing the second objective without reducing the quality of the first objective. Finally in the third step $\mathcal{A}_3$, we compute the solution optimizing the third objective without reducing the quality of the first and second objectives of the best solution found by the second step. In this section, we consider the real requirement of the SoICT which specifies that two reviewers of each student must be present in the jury of 5 members for his defense. Among two reviewers of each student, there is one who is inside the university and the other is outside the university.

2.1 Step 1

Clearly, the first objective function F does not depends on variables $xr[s], \forall s \in S$. Moreover, the variable $xr[s] (\forall s \in S)$ appear only in the constraint C4 for avoiding two defenses at a time slot are accommodated in the same room. This constraint can be replaced by an alternative using global constraint $AtMost$ specifying that there must be at most R defenses taking place at a given time slot. By using this alternative, for the first step, we can ignore the variable $xr[s] (\forall s \in S)$. For reducing the number of decision variables of the model, we fixed, for each student s, the values of $xp[s, 1]$ and $xp[s, 2]$ by his two reviewers. We denote $c_p[s, 1]$ and $c_p[s, 2]$ respectively the first and the second reviewers of the student s. Before describing the model, we initialize additional data structures from the input

- $f[p, d_j]$ measures the suitability of the professors and the subject domain d_j $(\forall p \in P, d_j \in SD)$
- $D^*[s] = S \setminus \{Sup[s]\} \setminus Rev[s], \forall s \in S$. $D^*[s]$ is the set of candidate professors for establishing the jury of the studeht s.
- $S^*[s]$ is the set of candidate time slots assigned to the defense of the student s. Note that a time slot t will

be a candidate if there are at least J professors having t as their available time slots and t belong to the available time slots of two reviewers of the student s. Hence, $S^*[s] = \{t \in SL \mid O[s,t] \geq J\} \cap T[Rev[c_p[s,1]] \cap T[Rev[c_p[s,1]]$ where $O[s,t] = \sharp\{p \in D^*[s] \mid t \in T[p]\}$, $\forall t \in SL, s \in S$.

- An array $o[1..|SL|]$ where $o[i] = |R|, \forall i \in \{1,\ldots,|SL|\}$

2.1.1 The Constraint Programming (CP) algorithm

The CP model consists of following components:

- Variables:
  - The decision variables of the model are now: $X = \cup_{s\in S}\{xt[s], xp[s,3], \ldots, xp[s,J]\}$. The domain of $xt[s]$: $Dom(xt[s]) = S^*[s]$, and the domain of $xp[s,j]$: $Dom(xp[s,j]) = D^*[s]$
  - For modeling the first objective function, we need auxiliary variables $xs[s,j], \forall s \in S, j \in \{3,\ldots,J\}$ where $xs[s,j] = f[d[s], xp[s,j]]$: $xs[s,j]$ measures the suitability in term of subject domain of the professor $xp[s,j]$ when he attends then defense of the student s.
- Constraints:
  - (CM1): $xp[s,j] < xp[s,j+1], \forall j \in \{3,\ldots,J-1\}, s \in S$
  - (CM2): $AtMost(xt[1],\ldots,xt[|S|], o)$
  - (CM3):
    * $xt[s_1] = xt[s_2] \Rightarrow xp[s_1,j_1] \neq xp[s_2,j_2], \forall j_1, j_2 \in \{3,\ldots,J\}$
    * $xt[s_1] = xt[s_2] \Rightarrow xp[s_1,j_1] \neq c_p[s_2,j_2], \forall j_1 \in \{3,\ldots,J\}, j_2 \in \{1,2\}$
  - (CM4): $\sum_{j=3}^{J} isHust[xp[s,j]] = J^{in} - 1$
  - (CM5): $BelongToSet(xt[s], xp[s,j], L), \forall s \in S, j \in \{3,\ldots,J\}$ where the global constraint $BelongToSet(x,y,S)$ states that $x \in S[y]$
  - (CM6): $BelongToSet(d[s], xp[s,j], D)$
  - (CM7): The constraint for deriving the value of $xs[s,j]$ from $xp[s,j]$

    $$xp[s,j] = p \Rightarrow xs[s,j] = f[p, d[s]]$$

    $\forall j \in \{3,\ldots,J\}, p \in P, s \in S$
- Objective function to be minimized

$$F(X) = \sum_{s\in S}\sum_{j=3}^{J} xs[s,j]$$

The search consists of instantiating values for variables of the model. The order of variables to be instantiated yielding different heuristics for the search.

2.1.2 The Constraint-Based Local Search (CBLS) algorithm

3. EXPERIMENTS

4. CONCLUSION