S32-63
137258
N93-18691
p. 5

Predit: a Temporal Predictive Framework for Scheduling Systems

E.Paolucci, E.Patriarca, M.Sem, G.Gini
Politecnico di Milano
Piazza Leonardo da Vinci 32
20133 Milano - Italy
Paolucci@ipmel1.polimi.it

Abstract

Scheduling can be formalized as a Constraint Satisfaction Problem (CSP). Within this framework activities belonging to a plan are interconnected via temporal constraints that account for slack among them. Temporal representation must include methods for constraints propagation and provide a logic for symbolic and numerical deductions.

In this paper we describe a support framework for opportunistic reasoning in constraint directed scheduling. In order to focus the attention of an incremental scheduler on critical problem aspects, some discrete temporal indexes are presented. They are also useful for the prediction of the degree of resources contention.

The predictive method expressed through our indexes can be seen as a Knowledge Source for an opportunistic scheduler with a blackboard architecture.

1. Formalization of scheduling problem and strategies for its solution

Scheduling can be formalized as a Constraint Satisfaction Problem (CSP)[Keng and Yun, 1989]. This approach is concerned with the assignment of values to variables subject to a set of constraints. In scheduling variables are constituted by activities start times and from resources allocation; for this reason we have to deal explicitely with two types of constraints: temporal relations among tasks and resources capacity [Fox, 86].

In our approach we assume that we have a set of plans to be scheduled, where a plan is defined as a partial ordering of activities. Each activity may require one or more resources and for each of them there can be alternative choices. Beside resources capacity can be used contemporarly by different tasks; for the sake of simplicity we will assume all resources with unary capacity.

Scheduling is an NP-hard problem and methods required for its solution must face this complexity. In our research we decided to focus our attention on contribution in scheduling coming from AI, and particularly on opportunistic reasoning [Hayes-Roth, 79].

We are concerned with the issue of how it's possible to focus the attention of an incremental scheduler on the most critical scheduling choices in order to evaluate which are the most critical points, which decisions seem to be the most promising in reducing search complexity and improving quality of resulting schedule.

Our strategy is to identify the most "solvable" aspects of the problem through the evaluation of the degree of interaction existing among activities belonging to different orders. The aim is to reduce the number of steps required to obtain a solution.

The necessity to overcome the limits of partial decomposition approach, such as order-based and resource-based decompositions, has led us towards an event-based perspective whit chronologically-grouped information.

This basic search strategy is realized through most-constrained and least-impact policies. Every step is divided into two parts: first the **most-constrained** policy selects dynamically on which agent must be focused scheduling attention; then, the **least-impact** policy chooses for that agent a value whose impact on the rest of the non-scheduled agents is as small as possibile. The goal is the identification of critical activities that heavily rely on the possession of highly contended temporal intervals or resources because of intra-order and inter-order interactions (**look-ahead strategy**).

This two policies need numeric indexes which, analyzing the particular structure of a problem, are able to measure the interaction among activities and resources in terms of **variable looseness** and **value goodness** [Sadeh and Fox, 88].

Variable looseness is the measure of how constrained is a resource or an activity; value goodness measures which variable value, among all the feasible ones, gives the least impact (i.e. a sort of maximum slack) on availability of feasible (and good) values for non-scheduled variables.

We identified some numeric indexes that contain information required to realize an event-based policy: these indexes are useful for different reasons:

1. This work has been partially supported by CNR PRF Project and Mauro grants

 C-3

- they make possible to point out critical resources and activities;
- they identify "island of certainty" that will be a part of problem solution;
- they give information about activities start times that have the least impact on non-scheduled activities.

This behaviour is a sort of "opportunistic reasoning" [Hayes-Roth, 79]: this term has been used to characterize a problem-solving process where reasoning is consistently directed towards those actions that appear most promising for solving a problem.

Our predictive approach, used together with an opportunistic reasoning, is also useful to detect unsatisfiable CSPs as soon as possible, simply by analyzing the indexes we defined. In this sense the system can be viewed as a Knowledge Source in a blackboard architecture, which assumes responsibility for preventive analysis of activities interactions and for the detection of prospective bottlenecks.

2. The predictive approach: basic assumptions

The main goal of our research was to provide a simple but complete inference mechanism to support scheduling, working in a discrete time domain. This mechanism is based on some indexes and is designed to perform an a-priori guidance for search in scheduling domain. We kept a particular attention on the efficiency and on the speed of such a mechanism, because we realized that such properties are necessary in scheduling systems for real applicative environments like, for instance, manufacturing ones. For this reason we decide to consider a discrete representation of time instead of a continous one.

Our indexes are based on the constraints analysis (and on the propagation of the temporal ones) and on a particular representation of existing time relations.

In terms of constraints analysis we differentiate between restrictions and preferences [Fox, 86]. Temporal preferences are represented through utility functions defined on activity start times that maps possible values onto utility levels ranging from 0 to 1. Moreover in our analysis we consider the existence of intra-order (among activities belonging to an order) and inter-order constraints (among activities belonging to different orders).

The model adoptated in representing time and in reasoning about temporal relation is based on the concept of lapse, that is defined as the period of time associated with an activity. In a temporal axis a lapse is represented by two temporal parameters, namely start time and end time. Relations between different lapses are expressed by two parameters [Paolucci, 90]:

- an **internal bound** (INT), which represents the minimun time interval which must separate the end of the first lapse from the begin of the second of two related lapses;
- an **external bound** (EXT), which represents the maximum time interval from the begin of the first lapse to the end of the second.

Through these two parameters it's possible to model any temporal relation in a scheduling problem. They are simpler than thirteen Allen's primitive relations; moreover, INT and EXT improve greatly the efficiency of numeric temporal reasoning, that is instead a limit in Allen's primitive.

3. The Predit indexes

Temporal relation constraints are used to describe partial orderings among activities as provided by the process planning step.

We will refer to the graph defined by these constraints, for a given CSP, as the CSP's Temporal Constraint Graphs (TCG).

We have to schedule a set of activities (A1,A2,...,AN). Let Ik be the time interval associated with Ak. Activities are connected by a set of temporal relation constraints, thereby forming a TCG. We view TCGs as undirected graphs. An Arc in a TCG indicates the presence of a temporal relation between two intervals represented by the couple INT-EXT (Fig.1).

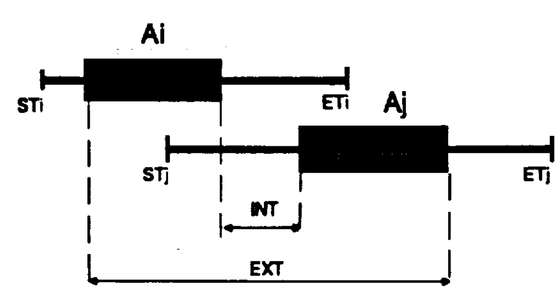

Figure 1

Additionally there are capacity constraints limiting the use of each resource to only one activity at a time. The next example presents a simple case of a TCG composed of two orders.

In order to provide a predictive support for opportunistic schedulers operating in a discrete time domain we have considered interactions among activities caused by temporal relations.

The first issue we faced was to detect as soon as possible during the scheduling the possible arising of conflicts due to interactions among activities.

For this issue we defined an index called Constraint Degree (CD), which measures the how tight is the link existing between two generic activities Ai and Aj connected by a temporal relation constraint (represented by INT-EXT) in a contraint graph.

Temporal relations between intervals may simply be expressed by using *potential inequalities* associated with the bounds of intervals such as :

[1] $D_i + D_j + INT \leq ET_j - ST_i$

[2] $D_i + D_j + INT \leq EXT$

The first inequality verifies that time interval composed of activities durations and Internal Bound is included in maximum temporal window allowed by Aj latest end time and Ai earliest start time.

The second inequality controls that the same time interval doesn't violate External Bound temporal constraint. These inequalities lead to define the CD formula through a multiplication of their members:

[3] $$CD_{ij} = \frac{(D_i + D_j + INT)^2}{EXT * (ET_j - ST_i)}$$

$$0 \leq CD_{ij} \leq 1$$

$D_k = A_k$ duration INT = internal bound between A_i and A_j

$ST_k = A_k$ start time EXT = external bound between A_i and A_j

$ET_k = A_k$ end time

The Constraint Degree is calculated on the notion of slack between two activities tied by temporal links.

- CDij = 1 means that Ai allows no slack to Aj (most constrained)
- CDij = 0 means that Ai allows maximum slack to Aj.

The CD computational algorithm considers all connected activities from the beginning to the end of the graph. Therefore, for ending activities we set CD index to zero (ending activities are not constrained, with temporal relation, with any other activity in the graph).

The validity of CD index is preserved by a previous optimization procedure in order to adjust activities temporal windows cutting out start time values that can never be involved in CD computation (the same is made for other indexes).

To sum up, the CD index detects (following the most-constrained policy) the most critical activities with respect to intra-order temporal relations (expressed by INT and EXT) and to temporal windows (expressed by activity start and end time).

The second index, called Preferential Start Time (PST), is a local measure of value goodness and globally, a measure of variable looseness for activities start times. It helps in choosing among all admissible start times the one that minimizes future conflicts. It is calculated between each pair of activities connected in the TCG (i.e. Ai and Aj) and it depends on the start time of the first activity (i.e. sti).

The main goal of PST index was to introduce some estimation rule for activity start times in order to identify the least impact values arising from intra-order interactions.

PST index is computed for every activity start time sti evaluated between earliest start time (STi), or value allowed by INT-EXT, and latest start time (ETi-Di), or value allowed by INT-EXT, increasing sti with the fixed time unit.

PST is expressed by the ratio:

[4] $$PST_{ij}(st_i) = \frac{int_{ij}(st_i)}{EXT - D_i - D_j - INT}$$

$$0 \leq PST_{ij}(st_i) \leq 1$$

where:

- $int_{ij}(st_i)$ = relative internal bound
- $EXT-D_i-D_j-INT$ = maximum slack between the activities

The numerator is calculated for st_i values from Earliest Start Time to the maximum allowed by temporal constraints, increasing each time st_i with a chosen time unit. It may be also considered as "actual" slack between the two activites corresponding to st_i value.

Therefore, the denominator may be viewed as the maximum slack between the two activities. The closer is PST_{ij} value to one, the greater is the slack between A_i and A_j. Therefore, for a generic activity PST measures for each admissible start time its goodness and likelihood to minimize future scheduling conflicts.

To compute activity Individual Demand for resources, we have combined the value goodness of every start time (expressed by PST) with the activities durations. Moreover, as assumed in [Sadeh-Fox, 88], an activity

A_i can use a resource R_j if A_i is active at time t and A_i uses R_j at time t to fulfill its resource requirement.

From each PST graph we achieve an Individual Demand graph (whose values are expressed by ID index) for each activity, expanding PST values with a lapse equal to the activity duration and adding all values in function of time. We obtain a histogram representing activity resource demand in function of time.

Individual Demand values are combined to measure resource Aggregate Demand (AD), always in function of time. AD shows when resource competition is particularly high and which are activities that heavily rely in the possession of these resources. AD values must be tightly evaluated in function of time because temporal constraint propagation doesn't allow for any resource preference (as explained before, we assume all resources with unary capacity). Therefore, AD index can estimate the amount of contention for each resource over temporal axis but only as a function of start time. Moreover, it's easy to improve this approach representing, for example, resources preferences with utility functions and propagating these resources reservations through the TCG graph.

Figure 2 shows a simple example in order to illustrate our graphic results concerning temporal discrete indexes presented above.

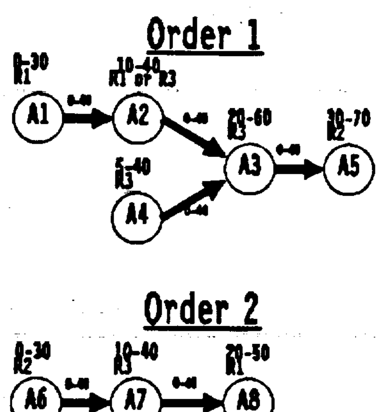

Figure 2

The temporal constraint associated at each linker is equal for all couple of activity and it is expressed by INT=0 and EXT=40. However, these values may be optimized as described before.

Start Time and End Time are expressed by numbers above each activity and the same is made for requested resources. For the sake of simplicity, in this example we have not introduced preferential start times (so activity start times are equally preferred).

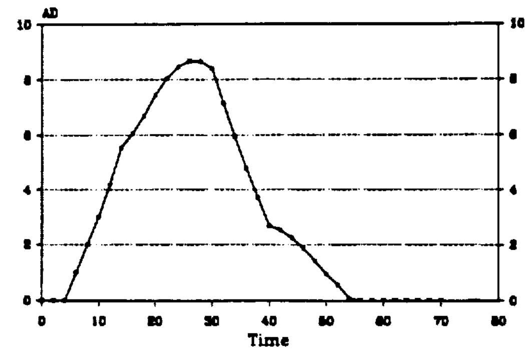

Figure 3

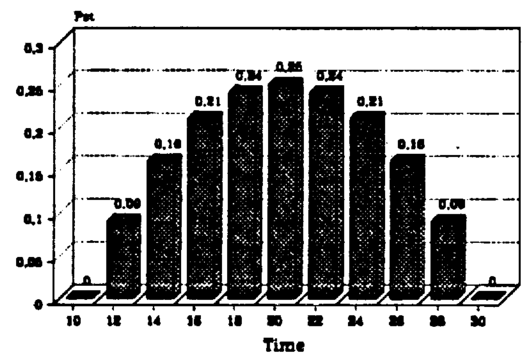

Figure 4

Next results are concerned with the reasonable steps that an Opportunistic Scheduler should achieve in order to produce the final Gantt chart.

A1	A2	A3	A4	A5
0,25	0,2	0,2	0.182	0

Table 1: CD values for order 1 activities

A6	A7	A8
0,25	0,25	0

Table 2: CD values for order 2 activities

Among the aggregate demands, the most highly contended resource is R3 (fig. 3), required by A2, A3, A4, A7; the next activities we will focus our attention on are A4, A3 and A7 (because A2 has an alternative in R1).

Taking a look at the CD indexes of order 1 and order 2, A7 appears to be the most constrained activity because of its highest CD value. Now, A7 PST graph (fig. 4) presents a maximum for t=20 and scheduling A7 with st=20 we can assign the resource R1 to the activity A2 at the same start time.

The same considerations based on temporal indexes evaluation allow the identification of other activity preferential start times leading to the Gantt chart presented below in fig. 5.

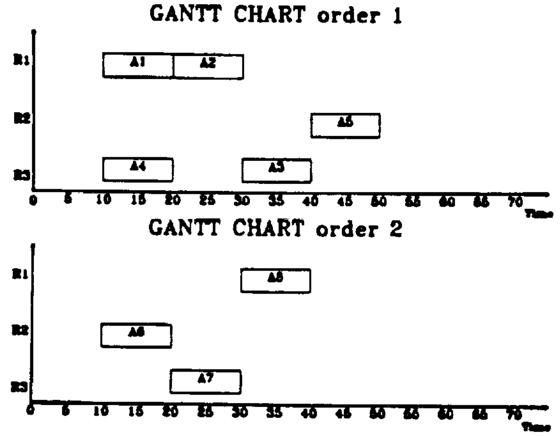

Figure 5

The quality of a schedule is based on the capability of the scheduler to satisfy a set of performance measures.

Moreover, a satisfiable schedule is always a compromise between the attempt to meet performance required and the necessity to respect all its constraints: schedule quality mirrors this trade-off. Each set of organizational constraints has its effects on final production schedules and, following the CSP formulation, if we change the constraints the solution will change too.

In order to improve schedule quality, our research is focusing on the evaluation of which impact might have an unexpected event on the resulting solution. PREDIT approach through the evaluation of discrete temporal indexes produces relatively accurate early predictions of activities behaviour as soon as PREDIT receives their changes and as long as constraints remain constant during indexes computation. The ability to react to changes that occur in dynamic environments providing a feasible solution in a sufficiently short time is very important expecially in manufacturing scheduling domain.

4. Concluding remarks

The approach we presented in this paper constitutes the basis for integrating an event-based mechanism and a predictive support in an opportunistic scheduling system.

We implemented this model in a MS-DOS environment with a particular attention towards speed performances. Our experiments indicate that our approach is successful in supporting opportunistic scheduling. This system is very efficient (it takes few seconds to calculate indexes in non-trivial real problems).

Our model seems to be highly appropriate for problems where the costs of backtracking is high because it's able to point out scheduling decisions that will minimize intra-order and inter-order conflicts. It increases significantly the performances of an opportunistic scheduler, making it possible to introduce such a tools in real applications.

Moreover the policies used by Predit to control the solution search (must constrained and least impact) can be used also in dynamic manufacturing environments. We are developing our research in this sense, also trying to support reactive scheduling and to manage multiagent production control systems.

References

[Fox, 1896], Fox M., Observation on the role of constraints in problem solving, *Proceedings Sixth Canadian Conference on AI, Montreal*, 1986.

[Hayes-Roth, 79], Hayes-Roth B. et al., Modeling planning as an incremental opportunistic problem, *Proceedings 3rd IJCAI, Tokio*, 1979.

[Keng and Yun, 89], Keng. N. and Yun D., A planning scheduling methodology for the constrained resource problem", *Proceedings IJCAI 1989*, pp. 998-1002.

[Paolucci, 90], Paolucci et. al., "CRONOS-III: requirements for a knowledge-based scheduling tool covering a broad class of production environments", in *Expert systems in engineering*, G.Gottlob and W.Nejdl (eds.), Springer-Verlag, 1990.

[Sadeh and Fox, 88], Sadeh N. and Fox M., Preference propagation in temporal capacity constraint graphs, *Technical report CMU-CS-88-193*.