Temporal Summarization of Plans

Karen L. Myers
AI Center, SRI International
333 Ravenswood Ave.
Menlo Park, California 94025
myers@ai.sri.com

Abstract

In many planning and scheduling applications, the challenge for human planners is not so much to find a solution to a problem, but rather to explore a range of options and understand the tradeoffs inherent to them. Effective search through complex solution spaces requires technology that can help a human decision maker understand the key aspects of candidate solutions. This paper presents an approach to summarizing temporal plans that focuses on identifying noteworthy temporal features. These techniques look for regularities or exceptional temporal elements, drawing on a modest domain theory to drive the search process. An evaluation of the method on a suite of MER mission plans illustrates the potential of the approach.

Introduction

In many planning and scheduling applications, the challenge is not so much to find a solution to a problem, but rather to explore a range of options and understand the tradeoffs inherent to them (Bresina et al. 2005; Myers & Lee 1999; Tate et al. 1998). Automated planning and scheduling techniques can be useful for these applications, but work to date in the AI community has focused primarily on the synthesis problem. Here, we consider providing automated support to assist with the exploration and understanding of the solution space in the form of tools to aid in summarizing temporal plans.

Our summarization techniques are domain independent and apply to a broad range of temporal plans. In this paper, however, we focus on plans generated using the MAPGEN mixed-initiative planning tool (Bresina et al. 2005) to support science planning for the Mars Exploration Robot (MER) mission. A MER plan consists of a set of activities to be performed by a rover in support of one or more science objectives. Typically, MER plans are created for a single Martian day (called a *sol*); however, science requirements can dictate the need for multi-sol plans.

Many potential plans are possible for a given set of science objectives; the challenge for planning in the MER domain is to identify actions that will best satisfy current science objectives and engineering constraints for a given sol. Some preferences are explicitly represented in the MAPGEN domain model; others are difficult to formalize and reside solely in the minds of the human planners. Ideally, a human planner would explore a range of options before committing to a particular solution. Because of time constraints, however, MER planners typically focus on an intended solution early on and have little opportunity to explore options. The work in this paper is motivated by the objective of making it easier for MER personnel to explore a broader range of solution options.

Our work on summarization of temporal plans focuses on techniques for identifying noteworthy temporal features. These techniques look for regularities or exceptional temporal elements in a plan, drawing on a modest domain theory to drive the search process. The techniques are intended to complement the plan summarization techniques described in (Myers 2006), which focus on structural aspects of a plan such as the tasks being performed, the strategies used to select them, and how resources are allocated. Furthermore, they are intended to augment rather than replace plan visualization capabilities, such as MAPGEN's timeline view of activities and resource levels.

We developed an implementation of our temporal summarization framework, called SITP (Summarization of Interesting Temporal Properties), and evaluated the framework on a suite of test plans from the MER mission. We believe that this evaluation shows that the summarization techniques can be helpful in increasing user understandability of complex temporal plans.

As with our prior work on plan summarization, domain-specific tools could be developed that provide more discriminating summarization capabilities. However, we feel that our domain-independent approach can provide significant value, particularly in the initial design stages of formulating a plan, when the user seeks to understand the high-level tradeoffs among alternative candidate solutions. In particular, our approach would be helpful to a human planner who wants a better understanding of the range of options available within a large, complex solution space. If desired, specialized evaluation tools could be applied to perform more expensive and time-consuming quantitative analyses to assess promising plans in more detail.

Our techniques can be applied to any plan represented in a machine-readable form, as they examine only the plan structure rather than the process by which the plan was generated. In particular, the plans could be authored manually, generated by automated planning tools, or developed within a mixed-initiative planning environment such as MAPGEN or PASSAT (Myers et al. 2002).

We begin by presenting a set of candidate relations for use in summarizing temporal aspects of plans. Next, we present a collection of methods to extract relevant relations from MER mission plans, followed by a summary of results in applying those methods to a test suite of MER plans. We close with a discussion of key properties of the approach, and a comparison to related work.

Temporal Relations

We adopt a model of plans similar to that employed in the MAPGEN system. More specifically, a plan consists of the following elements:

- a collection of *scheduled tasks*, each represented in terms of a task type, parameters, start time, duration, and a unique identifier
- *temporal constraints* over both task durations, and start and end times for an individual task or pair of tasks
- *temporal preferences* represented as weights on temporal constraints

In addition to elements of the plan, we make use of the notion of *events*, which correspond to temporal landmarks of various types. For the MAPGEN domain, for example, significant temporal events include the start and end times for a plan, as well as Martian sunrise and sunset.

Based on our plan model, we define a space of temporal relations that may be of interest when trying to understand a plan and assess its strengths and weaknesses. We split these relations into *local* properties, which correspond to relations that hold for an individual or small set of elements in the plan, and *global* properties, which apply to the plan as a whole.

The utility of individual instances of these relations for summarization and comparison of a temporal plan will depend highly on (a) the domain under consideration, and (b) context. The next section discusses the issue of utility for MER mission plans in more detail.

Local Properties

Task Duration Task duration is the amount of time that has been reserved for a task to run.

<Task1> has been scheduled to run for *N time units*

Such relations can be interesting when the time allocated is significantly smaller or larger than what is typical for tasks of that type.

Task Ordering Relative ordering information among tasks and events, in the form of the thirteen Allen relations (Allen 1984), could potentially be useful in understanding a plan. For example, consider the following templates:

- <Task1> *precedes/follows* <Task2>
  - *e.g.,* Visit Rock1 before Rock2
- <Task1> *precedes/follows* <Event1>
  - *e.g.,* Download data before end of Day-1
- <Task1> and <Task2> *start/end/occur* at the same time (i.e., synchronization)
- <Task1> and <Task2> *overlap*

A given plan will contain a large number of such relations, most of which will be of little significance to the user unless they are unusual or affect plan quality. For example, it is considered undesirable to have tasks overlapping with arm movements on the rover.

Task Placement Task placement characterizes where a task occurs within a plan. For example, task placement relations could be characterized by the template:

<Task> is *early/late/midway* in the plan

Task placements may be of interest when they are unusual. For example, communication between earth and the rovers usually occurs early or late during a sol; communicating mid-sol would be a deviation from standard practice.

Task Separation Task separation characterizes temporal gaps between a pair of tasks, or a task and an event.

<Task1> occurs *N time units before* <Task2>

Task separation relations may be of interest when they are either too short (e.g., insufficient buffer among actions to account for delays) or too long (e.g., too much idle time for the rover).

Temporal Preference Satisfaction Temporal preference satisfaction is a characterization of whether stated temporal preferences for positioning and duration of individual tasks, as well as intervals between tasks, are satisfied.

Global Properties

Task Spacing A number of global measures of temporal properties of plans and schedules related to task spacing can be found in the literature. One common global measure is the overall temporal duration of the plan, often called the "makespan" of the plan. Various measures of temporal robustness have also been proposed, including concepts such as temporal flexibility, temporal slack, and stability (for example, see (Policella et al. 2004)).

Other forms of task spacing may be of interest, depending on the nature of the application domain. For example, the following are relevant to assessment of plans in the MER domain:

- *Degree of overlap among tasks*. This value provides a measure of "density" of the plan. For many domains, it can be viewed as desirable to avoid extensive overlapping of tasks, which can complicate execution, monitoring, and repair.
- *Degree of synchronization among tasks*. Tasks are synchronized if there are temporal constraints that link the times of their relative execution. The dependencies in highly synchronized plans can contribute to decreased robustness and more complex execution.
- *Utilization/Fragmentation*. Utilization measures the degree of "busy" time, and hence can be useful in terms of understanding how effectively resources and time are being used. Fragmentation provides a related measure that characterizes occurrences and distributions of idle time. In the MER mission, for example, a measure of fragmentation can be helpful in identifying significant periods of rover downtime.

Task Frequency A task can be unusual either because it occurs (e.g., rover deep sleep), occurs more times than normal (e.g., a large number of imaging operations), or does not occur (e.g., no imaging). The occurrence or nonoccurrence of tasks is in general not a temporal feature of a plan. For domains such as science mission planning, however, in which the objective is to maximize productivity of available resources by scheduling the most extensive possible set of activities from an overconstrained set, the nonoccurrence of tasks can often be attributed to temporal conflicts.

Time Allocation Just as the time devoted to a particular task can be of interest, so can the total amount of time allocated to tasks of a particular type. Time allocations can be of interest both because they are unusually low (e.g., only 5 minutes of data download) or unusually high (e.g., awareness of extensive heating operations is important because they can draw down battery power significantly).

Temporal Patterns Repeated local temporal relations of the type discussed above can reveal important structural characteristics of a plan. For example, it could be valuable to recognize that all tasks of a certain type are planned before all tasks of a different type (e.g., all pictures are taken before any traversals are made), or that all tasks occur before/after some significant event (e.g., all imaging is completed by noon).

Global Temporal Preference Satisfaction The aggregate degree of satisfaction of temporal preferences will generally be an important factor in measuring the overall quality of a plan.

Summarization Methods

As noted above, any of the relations described in the previous section could be of potential interest to a user, depending on the circumstances. Our approach to identifying interesting temporal features of plans relies on the use of background information for a domain to help with the assessment of the significance of a particular relation for a given plan.

In developing our summarization framework for MAPGEN, we focused on the local relations of task duration and task ordering, and the global relations of time allocation, task frequency, and temporal patterns. Task placement relations were considered only relative to temporal landmarks (see the Task Ordering Relations section); other forms of task positioning are not important for MER plans.

Temporal preference satisfaction, both local and global, is extremely important for assessing overall plan quality. However, for weighted preferences, it is straightforward to identify those that are most significant within a given plan, and schemes for aggregating preferences over temporal plans are well studied (Khatib et al. 2001).

Task spacing relations, although not considered here, could potentially provide insight into important temporal features of a MER plan. We note, however, that MAPGEN includes a user interface that displays timelines for actions and expected resource levels. In effect, this interface provides a direct visualization for the user of task overlap, synchronization, and utilization/fragmentation.

The rest of this section discusses the specific temporal feature summarization methods that we developed for the MER domain. These methods were implemented in a system called SITP (Summarization of Interesting Temporal Properties) that was evaluated on a test suite of ten sample plans. The results of the evaluation are described in the following section.

Task Duration

For task duration, we are interested in identifying tasks that have been scheduled for significantly shorter or longer periods of time than is typical. The main challenge here is defining what constitutes a typical duration for a task.

One simple approach is to define an interval (or intervals) of typical values *a priori*, drawing on inputs from domain experts. While straightforward, this approach imposes additional knowledge engineering responsibilities; furthermore, depending on the source of the knowledge and how the models are elicited, the models may have varying degrees of accuracy.

A preferred approach would be to learn the interval from historical data by applying data mining techniques. In statistical terms, the values of interest are those that can be viewed as *outliers* for a distribution (Moore and McCabe 1999). While a variety of definitions for outliers have been proposed in the literature, many require identification of a distribution for the population of values, which in itself is a difficult task (Knorr and Ng 1997). We take a simpler, commonly used approach grounded in the notion of the *interquartile range*.

Definition 1 [Interquartile Range]. Let S be a population of values with first quartile Q_1 and third quartile Q_3. The interquartile range for S is defined to be $|Q_3 - Q_1|$.

Definition 2 [Typicality Interval, Outlier]. Let S be a population of values with interquartile range R, first quartile Q_1 and third quartile Q_3. The typicality interval for S is defined to be $[Q_1 - 1.5*R, Q_3 + 1.5*R]$. A value v is an *outlier for S* iff v lies outside the typicality interval for S.

For the experimental results reported in this paper, the lower and upper bounds for task duration were set to the bounds on the typicality interval computed for the population of values in the test suite of plans. As those results show, this definition was effective in identifying small numbers of cases that stood out from the population. We analyzed plots for all the task durations and found that in all but one case, the definition matched our intuitions for what we would consider interesting deviations from the population of values. In the one exceptional case, the most natural cutoff for identifying outliers was ambiguous.

Task Ordering

The most important task ordering relations for the MER domain are those that link certain types of tasks to key temporal landmarks for sols. These landmarks consist of *sunrise*, *sunset*, *plan-start*, and *plan-end*. Another useful temporal landmark is *noon*, when a rover would have maximal exposure to the sun, and hence have maximal potential for recharging its battery.

Communication was identified as one class of task that can often be anchored to temporal landmarks (specifically, tasks of type X_CARRIER_LGA and X_DFE_HGA). For example, a typical sol begins with an uplink of commands to the rover, and near the end of the sol, the rover downloads data collected during that sol.

In the SITP system, we provided a framework in which to declare relative time intervals within a sol during which tasks of a designated type would be expected to occur. To handle expectations about communication tasks, we defined the additional temporal landmarks *end-of-early* and *start-of-late*. *End-of-early* represents the transition from early in the day and was defined on a per-sol basis as commencing at two hours past *plan-start*. Similarly, *start-of-late* marks the transition to end of the day and was defined on a per-sol basis as commencing two hours prior to *plan-end*. Task types can have associated *time-of-occurrence* properties, which indicate when those tasks should appear relative to temporal landmarks. For example, the communication events of type X_CARRIER_LGA and X_DFE_HGA were declared to occur either early or late in the day.

Task Allocation

For task allocation, we were interested in identifying types of tasks to which significantly more or less than the typical time had been allocated within a plan. We adopted an approach similar to that used with task duration, relying on the definition of an outlier grounded in the interquartile range to identify interesting cases. The experimental results also employed allocation bounds computed from the distributions of values in the test suite of plans, and produced intuitive results.

Task Frequency

SITP supports the declaration of a *frequency* property for tasks designed to characterize how often they occur. In general, a range of different models of frequency could be useful in different settings. For example, in a personal calendar, one could imagine that a typical class of meetings takes place weekly, biweekly, or monthly, while other tasks (exercise, lunch) may occur daily.

Originally, we had expected to identify a set of MER tasks that occur rarely, and hence would be worth noting in the summarization of a plan. As our knowledge of the domain increased, we determined that although it may be unusual to perform certain MER activities in particular situations, no individual activities were inherently rare. It is the case, however, that certain tasks are expected to be performed within every sol or every plan, and the absence of one of these tasks is worthy of note.

For these reasons, we chose to focus on frequency values drawn from the set {*hourly*, *per-plan, per-sol*} for the MER domain. Alternatively, it could also be useful to consider a qualitative set of values, such as {*low*, *medium, high*}, to characterize the rate of occurrence in a more general manner.

Temporal Patterns

We explored the occurrence of temporal patterns defined over certain task types in an effort to find cases in which all (or most) instances of a task within a plan exhibited some potentially noteworthy property. In particular, SITP searches for patterns of unexpected task durations and task ordering relations, which we call the *property of interest* for the pattern.

More specifically, for a given plan, SITP identifies patterns in which a given task type occurs some minimum number of times (*required-#occurrences*) and for which some percentage of the instances of that type within the plan (*required%*) satisfy a designated property of interest. The experiments described later used the values *required-#occurrences*=3 and *required%*=0.7.

Observation Significance

We use the term *observation* to refer to a temporal relation within a plan that has been determined to be potentially noteworthy using the methods described in the previous section. Because some of the observed temporal relations will have greater import than others, we associate *significance measures* with observations. Significance measures are normalized to lie in the interval [0,1] and are defined as the product of two factors: (1) a measure of the importance of the tasks associated with an observation, and (2) a *strength of occurrence* value.

One natural basis for defining a measure of importance is task *priority*. MAPGEN associates priorities with tasks on a scale that ranges from 0 (maximum priority) to 5 (minimum priority). For the test results presented below, these priorities were normalized to lie in the range [0,1] via the following computation:

(/ (+ 1 (abs (- *TaskPriority MinPriority*)))
(+1 (abs (- *MaxPriority MinPriority*))))

The priority associated with local observations was set to this normalized task priority; the priority for global observations was set uniformly to *MaxPriority.* For domains without task priorities, background knowledge about the general importance of task types could be used instead.

The strength of occurrence values are designed to reflect the degree to which the observed relation holds. We define these values in a relation-specific way, as described below.

```
# Plan Steps: 63
Start time: 2004-08-19T03:40:15      End time: 2004-08-20T04:06:30

:MISSING-DAILY-TASK (1.0): X_CARRIER_LGA
:LONG-DURATION-TASK (1.0): MTES_SPECTRA_OR_INTERFEROGRAMS at 2004-08-19T04:38:46 (duration
   35:12; upbound 13:32)
:LONG-DURATION-TASK (1.0): PANCAM_SINGLE_POSITION at 2004-08-19T04:31:32 (duration 7:14;
   upbound 3:30)
:LONG-DURATION-TASK (.83): PANCAM_SINGLE_POSITION at 2004-08-19T04:22:08 (duration 9:24;
   upbound 3:30)
:LONG-DURATION-TASK (.48): CPU_ON at 2004-08-19T03:21:03 (duration 2:45:32; upbound
   1:44:54)
:UNDERALLOCATION (.48): UHF_COMM given 29:00 but has bound 55:30
:LONG-DURATION-TASK (.19): NAVCAM_MOSAIC at 2004-08-20T03:49:53 (duration 4:16; upbound
   3:36)
:LONG-DURATION-TASK (.02): X_DFE_HGA at 2004-08-20T03:50:14 (duration 31:00; upbound 29:00)
:LONG-DURATION-TASK (.01): X_DFE_HGA at 2004-08-19T03:27:04 (duration 30:00; upbound 29:00)
```

Figure 1. SITP Output for Test Plan 223

For tasks that were to occur within a given time interval, the strength of occurrence is defined to be the distance to the closest boundary of the interval, divided by the size of the interval, provided the value is no greater than 1; otherwise, the value is capped at 1. For situations in which the task could appear in multiple intervals, the strength of occurrence is defined relative to the closest boundary over all the possible intervals, divided by the size of the closest interval.

For tasks whose durations are determined to be unusually short, the strength of occurrence is defined to be the distance to the minimum duration associated with that task, divided by that minimum duration, provided the value is no greater than 1; otherwise the value is capped at 1. Similarly, for tasks with unusually long duration, the strength of occurrence is defined to be the distance to the maximum duration associated with the task, divided by that maximum duration, with the value capped at 1.

For temporal patterns, strength of occurrence is assessed as a measure of the ratio of the number of tasks of a given type that satisfy the property in question to the total number of tasks of that type in the plan.

For tasks that are noteworthy, because of either their inclusion or their omission, we use a uniform strength of occurrence of 1.

Evaluation on MER Mission Plans

We evaluated our temporal summarization framework by applying it to a suite of ten test plans provided by the MAPGEN team. The test plans are based on actual MER mission plans created via MAPGEN, but massaged slightly to eliminate the need for specialized resource reasoning components not relevant to our work. Each plan was generated for a different mission day and so addressed different overall science objectives. Plans are referred to by the number of the sol for which they were created (i.e., plan 223 was created for the 223rd sol of the mission).

Examples

Figures 1 and 2 show the results produced by SITP on two plans from the test suite. For completeness, we include all observations produced by SITP; in actual usage, SITP outputs would be filtered to account for user preferences, observation strength thresholds, and total number of observations. In the output, durations are represented in the format "*hh:mm:ss*", while dates have the form *"yyyy-mm-ddThh:mm:ss"*.

Most observations begin with a prefix of the form

<*observation-type*> (<*significance*>):

where <*significance*> denotes the significance measure defined earlier. (Temporal patterns present an exception and are discussed separately at the end of this subsection.) Each observation prefix is followed by a more detailed characterization of the observation, whose syntax depends on the observation type, as defined below.

Observations related to duration have type `:SHORT-DURATION-TASK` or `:LONG-DURATION-TASK`, indicating that the task was shorter or longer than expected. The remainder of the observation is structured as either

<*task-type*><*timestamp*> `(duration` <*duration*>`; lowbound` <*bound*>`)`

for `:SHORT-DURATION-TASK` observations or

<*task-type*><*timestamp*> `(duration` <*duration*>`; upbound` <*bound*>`)`

```
# Plan Steps: 39
Start time: 2004-12-01T21:23:39      End time: 2004-12-03T00:07:22
:LONG-DURATION-TASK (1.0): NAVCAM_MOSAIC at 2004-12-02T03:07:56 (duration 11:42; upbound
   3:36)
:LONG-DURATION-TASK (1.0): PANCAM_SINGLE_POSITION at 2004-12-01T23:47:30 (duration 7:14;
   upbound 3:30)
:LONG-DURATION-TASK (1.0): PANCAM_SINGLE_POSITION at 2004-12-01T23:40:16 (duration 7:14;
   up bound 3:30)
:LONG-DURATION-TASK (.35): CPU_ON at 2004-12-02T01:39:39 (duration 2:28:44; upbound
   1:44:54)
:SHORT-DURATION-TASK (.13): PANCAM_SINGLE_POSITION at 2004-12-02T23:50:39 (duration 2:03;
   lowbound 2:22)
:SHORT-DURATION-TASK (.13): PANCAM_SINGLE_POSITION at 2004-12-02T23:40:51 (duration 2:03;
   lowbound 2:22)
:SHOULD-BE-EARLY-OR-LATE (.05): X_CARRIER_LGA at 2004-12-01T23:30:23 (06:44
   beyond early window, 16:36:59 before late window)
```

Figure 2. SITP Output for Test Plan 325

for `:LONG-DURATION-TASK` observations.

Observations related to task ordering have type `:SHOULD-BE-EARLY`, `:SHOULD-BE-LATE`, or `:SHOULD-BE-EARLY-OR-LATE`, indicating that the task started outside of the defined early, late, or early or late windows, respectively. The remainder of the observation is structured as

<task-type><timestamp> `(`*<duration>* `after early window)`

for a `:SHOULD-BE-EARLY` observation,

<task-type><timestamp> `(`*<duration>* `before late window)`

for a `:SHOULD-BE-LATE` observation, and

<task-type><timestamp> `(`*<duration1>* `after early window;` *<duration2>* `before late window)`

for a `:SHOULD-BE-EARLY-OR-LATE` observation.

Observations related to missing tasks have the form

`:MISSING-DAILY-TASK (`*<significance>*`):` *<task-type>*

Observations related to over- or underallocation of time to a given task type are represented by the observation types `:OVERALLOCATION` and `:UNDERALLOCATION`, and are presented as follows:

<task-type> `given` *<allocation>* `but has bound` *<bound>*

Temporal pattern observations have the observation type `:ALLTASKS` and are presented as

`:ALLTASKS` *<relation-type>* `(`*<significance>*`):` *<task-type>* `(`*<n>* `occurrences)`

Here, *<relation-type>* specifies the type of temporal relation that constitutes the property of interest over which the pattern is defined; *<n>* denotes the number of instances of the task type in the plan that satisfied the relation.

Analysis of Results

The total number of observations produced for the ten plans in the test suite was 76, with the number for individual plans ranging from 3 to 15. The median number of observations was 7.

Interestingly, there was no correlation between plan size (i.e., number of tasks) and number of observations extracted by SITP, with one exception (see Figure 3). The plan sol-300 was the largest plan (87 tasks) and produced the most observations (15). Closer examination of plan sol-300 shows that the reason for the high number of observations is not the size of the plan, but rather its temporal scope: while the other plans in the test suite occur within a single sol, sol-300 spans three sols. Some aspects of our background model of typicality were not well suited to multi-sol plans. So, for instance, the longer temporal scope of the plan resulted in more communication tasks as well as increased usage of the CPU than was typical in single-sol plans. These led to SITP producing `:OVERALLOCATION` observations for tasks of type UHF_COMM, X_CARRIER_LGA, X_DFE_HGA, and CPU_ON. Furthermore, the expectation that communications would occur early or late in the plan broke down, as reflected in two observations of type `:SHOULD- BE-EARLY-OR-LATE` each for the task types X_CARRIER_LGA and X_DFE_HGA.

At one level, the abundance of observations for the plan sol-300 can be used as a justification for our approach: the multi-sol plan is unusual and its atypicality comes through in the temporal summaries. This case does, however, point

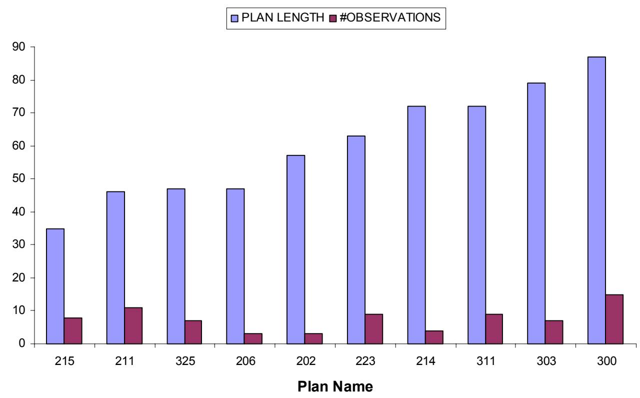

Figure 3. Plan length and number of observations for plans in the MER test suite

out the need to contextualize the domain models to a greater extent – a point to which we return in the Discussion section.

The most frequent observation type was `:LONG-DURATION-TASK` (54 occurrences) followed by `:SHORT-DURATION-TASK` (7 occurrences). The asymmetry between tasks that were atypically long or short occurred because populations of task durations in the sample plans resemble exponential distributions, with a preponderance of short durations along with a tail containing some small number of longer durations.

Only a single `:SHOULD-BE-EARLY-OR-LATE` observation occurred outside of plan sol-300, and with a weak strength of occurrence (.06). Two additional `:OVERALLOCATION` observations were produced outside of plan sol-300, one for UHF_COM tasks, another for MTES_SPECTRA_OR_INTERFEROGRAMS tasks. Two plans lacked any X_CARRIER_LGA task, which was declared in the background model as occurring daily. Finally, a single `:UNDERALLOCATION` observation was produced (for the UHF_COM task type).

No temporal patterns were detected, which indicates that the individual plans did not exhibit high degrees of variability in unexpected task durations and task ordering relations (which were the two properties of interest considered for temporal patterns).

Overall, we believe that the results on the test suite show that our techniques can be effective in highlighting potentially significant temporal elements of plans. As such, they should make it easier and faster for a user to explore a rich space of potential solutions.

Discussion

Our approach to summarizing interesting temporal features of plans has the advantage that it is straightforward to implement and simple to compute. Although we have emphasized summarization for *plans*, our techniques apply equally well to *execution traces* for plans. In particular, these methods could be useful in spotting, after the fact, unusual aspects of what actually happened during execution, as opposed to what was planned for execution.

One potential issue with our approach is the effort required to formulate the background model used to drive the identification of interesting temporal features. The model presented here assumes uniformity of temporal characteristics over all plans. For the MER domain, however, time allocations, durations, and priorities can vary with the type of sol being planned (e.g., Drive, Approach, Remote Science, Scratch-and-Sniff) as well as the temporal scope of the plan (e.g., single- vs multi-sol). This specialization introduces the need to contextualize the models in various ways. If the number of contexts is relatively small, as is the case for MER plans, this contextualization should not lead to a significant increase in the effort required to build the model.

Although we have focused on extracting summaries of temporal features for an individual plan, it would be straightforward to develop a plan comparison tool built on SITP that could identify commonalities or differences in temporal features across plans. A comparison tool of this type would be valuable in aiding with exploration of a complex solution space. It could also help a user understand the effects of possible or actual changes to be made to a plan.

Individual users will most likely vary in their perspectives on what is important or noteworthy in a plan. If put into operational use, a tool like SITP should be augmented with a *customization* capability grounded in a user profile that designates the types of temporal relations and tasks of greatest interest to the user. This profile could take the form of a series of weights defined for the relations and tasks, and would be specified a priori on a per-user basis. However, it could potentially be adapted over time by incorporating feedback from the user on the types of observations that proved to be helpful in practice.

For certain types of domains, it may be important to provide some form of summarization of the overall temporal structure or phasing of a plan in addition to the observations extracted by SITP. For example, a structural summarization would be helpful for complex military plans, in which the overall temporal flow and temporal synchronization are key to understanding plan strategy. For MER missions, such an overall temporal summarization is not important given that plans consist of a collection of tasks with only loose temporal coupling. Similarly, a collection of calendar entries would have no global strategy, but rather only limited intertask linkage.

Related Work

Relatively little work has been published to date on summarizing plans. Most of what can be found in the literature focuses on plans without explicit temporal representations (e.g., (Mellish and Evans 1989; Myers 2006; Young 1999)). (Clement et al. 2007) describe a method for generating hierarchical abstractions of state conditions and resource usage for temporal plans.

However, the motivation for that work is coordinating distributed schedulers rather than improved user understandability.

There has been work within the constraint reasoning community on the related topic of explanation, most of which focuses on explaining temporal conflicts that arise during the search for a solution. As pointed out by (Smith et al. 2005), most of this work assumes a *system-level* perspective in that explanations are grounded in constraints and reasoning processes of the system rather than at a level that targets a user's conceptualization of the domain.

Some efforts, however, adopt a more *user-centric* view of explanation. (Smith et al. 2005) introduce a rich ontology for grounding both explanations of temporal constraint conflicts and recommendations for repairing them. (Bresina and Morris 2006) also define techniques to explain and resolve temporal inconsistencies, within the context of MAPGEN plans. Their work focuses on distilling from a complex temporal no-good, which may have hundreds of constituent constraints, a small set that communicates the essence of the conflict.

Conclusions

We have described an approach to summarizing temporal plans that identifies regularities or exceptional temporal properties, drawing on a modest domain theory to drive the search process. This method can be used to highlight aspects of a plan that could potentially impact a user's willingness to accept that plan. We developed an implementation of the framework, called SITP, and evaluated the framework on a suite of test plans from the MER mission. This evaluation shows promise that the summarization techniques can be helpful in increasing user understandability of complex temporal plans.

In future work, we will seek validation of our summarization approach from members of the MAPGEN community. We will also explore the application of our framework to other types of temporal plans, most notably personal calendars.

Acknowledgments. This work was supported by NASA Grant NNA-04CL08A. The author thanks John Bresina for his many insights into the MER domain, as well as for providing the MAPGEN software and suite of test plans. The author also thanks John, Lina Khatib, Paul Morris, Conor McGann, Andres Rodriguez, and Michael Wolverton for valuable feedback on the work.

References

Allen, J. 1984. Towards a General Theory of Action and Time. *Artificial Intelligence*, 23(2), 123-154.

Bresina, J., Jónsson, A., Morris, P., and Rajan, K. 2005. Activity Planning for the Mars Exploration Rovers, *Fourteenth International Conference on Automated Planning and Scheduling*, Monterey, CA, 40-49.

Bresina, J. and Morris, P. 2006. Explanations and Recommendations for Temporal Inconsistencies, *Proceedings of the Fifth International Workshop on Planning and Scheduling for Space*.

Clement, B., Durfee, E., and Barrett, A. 2007. Abstract Reasoning for Planning and Coordination. *Journal of AI Research*, Vol. 28, 453-515.

Khatib, L., Morris, P., Morris, R. and Rossi, F. 2001. Temporal Constraint Reasoning with Preferences. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 322-327.

Knorr, E. M. and Ng, R. T. 1997. A Unified Notion of Outliers: Properties and Computation. In *Proceedings of Knowledge Discovery and Data Mining*, 219-222.

Mellish, C. and Evans, R. 1989. Natural Language Generation from Plans. *Computational Linguistics*, 15(4).

Moore, D. S. and McCabe G. P. 1999. *Introduction to the Practice of Statistics.* New York: W. H. Freeman.

Myers, K. L. and Lee, T. J. 1999. Generating Qualitatively Different Plans through Metatheoretic Biases. *In Proceedings of AAAI-99*, AAAI Press, Menlo Park, CA.

Myers, K. L., Tyson, W. M., Wolverton, M. J., Jarvis, P. A., Lee, T. J., and desJardins, M. 2002. PASSAT: A User-centric Planning Framework. In *Proceedings of the 3rd Intl. NASA Workshop on Planning and Scheduling for Space*, Houston, TX.

Myers, K. L. Metatheoretic Plan Summarization and Comparison. 2006. In *Proceedings of the 16th International Conference on Automated Planning and Scheduling*, Windermere, Cumbria, U.K.

Policella, N., Smith, S.F., Cesta, A., Oddi, A. 2004. Generating Robust Schedules through Temporal Flexibility. In *Proceedings of the 14th International Conference on Automated Planning and Scheduling*, Whistler, Canada.

Smith, S. F., Cortellessa, G., Hildum, D. W., and Ohler, C. M. 2005. Using a Scheduling Domain Ontology to Compute User-oriented Explanations. In *Planning, Scheduling and Constraint Satisfaction: From Theory to Practice* (ed., L. Castillo, D. Borrajo, M.A. Salido, A. Oddi), IOS Press.

Tate, A., Dalton, J., and Levine, J. 1998. Generation of Multiple Qualitatively Different Plan Options. In Proceedings of Artificial Intelligence Planning Systems, Pittsburgh, PA.

Young, R. M. 1999. Using Grice's Maxim of Quantity to Select the Content of Plan Descriptions, A*rtificial Intelligence*, 115(2).