4D Heat Maps: Visualizing Uncertain Resource Utilization Over Time

Robert Gove*
Booz Allen Hamilton

Brandon Herzog†
Booz Allen Hamilton

Abstract

Understanding the utilization of many resources over time can be difficult, especially in the face of uncertainty. Often there are four variables to visualize: Resource name, time periods, predicted utilization amount, and uncertainty. (This is in contrast to tasks, which are not repeated in the schedule and do not have utilization amounts.) It can be difficult to visualize an overview of all variables simultaneously using existing techniques such as heat maps, small multiples, or resource charts. Common heat maps encode three visual variables (color and horizontal and vertical position of each cell), but we propose using augmented heat maps to visualize the resource name, time period, predicted utilization amount, and uncertainty using the vertical position, horizontal position, color, and size of each cell, respectively. This visualization essentially provides a summary of histograms over time. We use this 4-Dimensional heat map to visualize resource utilization in the output from a Monte Carlo simulation of an anonymized government schedule.

1 Introduction

Understanding resource utilization is an important part of many scheduling problems, from scheduling hospital beds [1] to scheduling employees on projects. (*Utilization* refers to the amount of a resource that is used at a particular point in time, e.g. if half of a software engineer's time is being used, then the software engineer is 50% utilized.) This task is often complicated because there may be dozens of resources to track and there may be uncertainty about each resource's utilization over time, and this task is not well supported in existing project scheduling tools. Organizations often suffer from cost and/or schedule overrun [4], which could be caused by inefficiently scheduling their resources. Analysts and project managers often need to answer complex questions about resource utilization. For example, if tasks take longer than expected to finish, will that cause a resource shortage? Or, when will resources be available to work on other projects?

Utilization predictions can be uncertain for several reasons. For example, a task can take longer to finish than originally expected, which may cause a resource (such as an engineer) to be scheduled on several tasks simultaneously, or a task may need more resources than expected (e.g. a full-time engineer instead of part-time). This uncertainty can be modeled as probability distributions and analyzed using Monte Carlo simulation.

Resources can be visualized in resource charts, similar to Gantt charts, but they do not display uncertainty in how much or for how long the resource will be utilized. In this paper we present a new way to visualize uncertain resource utilization using a 4-Dimensional heat map. Similar 4D heat maps have been developed before [3], but to our knowledge they have only been used for microarray data, and we have not seen an example data analysis. We discuss our design decisions, and demonstrate the 4D heat map using the resources in an anonymized government project schedule. We conclude with lessons learned and potential enhancements.

*e-mail: govejr_robert@bah.com
†e-mail: herzog_brandon@bah.com

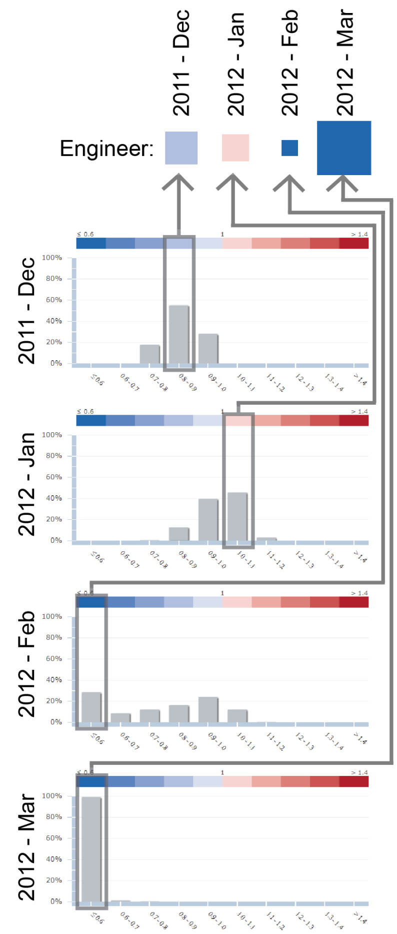

Figure 1: The process of creating a 4D heat map row visualizing uncertain utilization of an engineer over four time periods.

2 System Design and Data Analysis

We expanded Polaris, our commercial Monte Carlo simulation project analysis tool, to include to include the resource utilization visualization described below. The tool allows users to build a model of their schedule uncertainty by applying probability distributions to task duration and resource utilization rates. The tool then performs a Monte Carlo simulation with 1,000 or more trials, and users can perform statistical analyses on the results. Because of the probability distributions in the schedule model, tasks' durations can extend or decrease, tasks can require more or less resource utilization, or both.

To visualize resource utilization predictions and uncertainty, we created a heat map where each row is a resource and each column is a time period (months, quarters, or years). We color cells in the heat map based on the predicted resource utilization rate. We use color because, unlike size, color can indicate if a value is above or below a threshold. We use a diverging color scale where dark- to light-blue colors indicate high- to low-degrees of underutilization, and dark- to light-red colors indicate high- to low-degrees of overutilization. (E.g. if one resource is available and the prediction is to utilize half

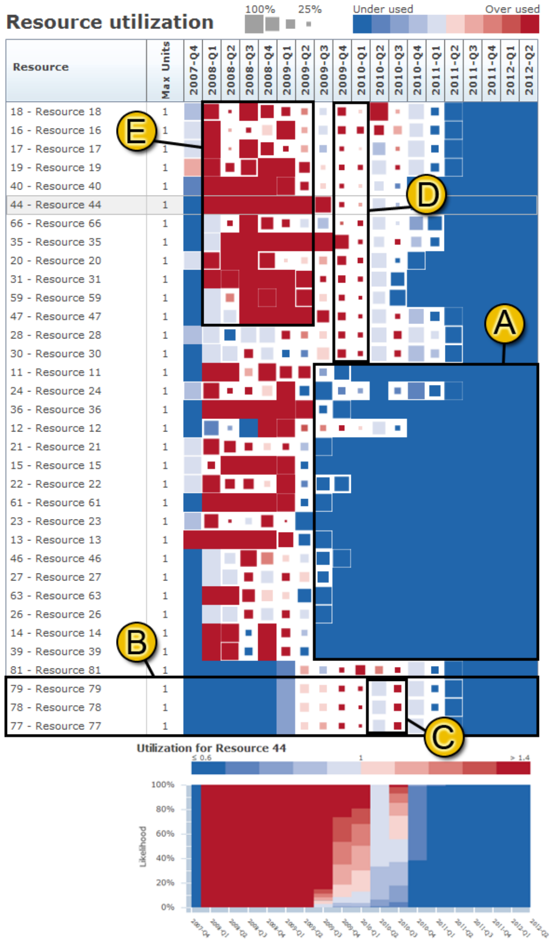

Figure 2: TOP: Resource utilization heat map, sorted by similarity to Resource 18. (A) Half of resources have availability beginning 2009-Q3. (B) Several resources have similar utilization patterns (e.g. 77, 78, 79). (C) Unexpected change from confident slight underutilization to unconfident high overutilization. Further investigation of the tasks utilizing these resources revealed two sets of missing schedule dependencies. (D) Many resources have uncertain utilization in the middle of the schedule. (E) Many resources are overutilized in the beginning. BOTTOM: Utilization area chart for Resource 44. Further investigation revealed that 2010-Q3 uncertainty was caused by many uncertain tasks utilizing small amounts of Resource 44.

of the resource, the color will be dark blue.) We chose a blue-red color scale over the more popular green-red color scale to reduce issues with color blindness.

To visualize the degree of uncertainty in the prediction, we encode the prediction likelihood as a fourth variable using the size of the heat map cell. We chose to encode likelihood as size because high likelihood results in a larger cell with more visual weight, whereas low likelihood results in a small cell with less visual weight indicating that users should click the cell to see more details.

See Figure 1 for an example of constructing heat map row cells from time period resource utilization histograms. We build a histogram for each time period (Figure 1 bottom), where the horizontal axis represents resource utilization bins, and the vertical axis represents percent of the simulation outcomes that fall in a given bin. In this example, we are computing the utilization for one engineer over four time periods. For each histogram, the bin with the highest percentage of outcomes (tallest bin) is chosen as the predicted resource utilization, thus determining the heat map cell color (Figure 1 top). We use the percent of simulation outcomes in the chosen bin (height of the bin) to size the heat map cell; consequently, a shorter histogram bar means a smaller heat map cell.

Our heat map rows can be sorted in three ways: Sorting by utilization in a specified time period, sorting by utilization across all time periods, and sorting by similarity to a specified resource. These sorting methods allow users to find (1) over- or underutilized resources in a given time period, (2) resources that are over- or underutilized across time periods, and (3) resources that have similar utilization patterns over time. See Figure 2 for an example visualization. We chose to use sorting instead of a clustering because runtime performance is critical in our software, and clustering algorithms that perform leaf ordering for perceptual ordering are noted to cause performance issues [5] [2].

Users can click a resource name to see a stacked area chart of the resource's utilization over time (Figure 2 bottom). This chart uses the same color scheme as the heat map. Time extends horizontally, allowing users to see the resource's utilization possibilities evolve over time. Users can click on an individual cell in the heat map, or on a time period in the area chart, to see the histogram of the resource's utilization in that time period (e.g. in Figure 1). Polaris can then show all tasks that utilize the resource at any point in the schedule (if a resource name is selected) or in only the selected time period (if an individual cell is selected).

For an example analysis, we added probability distributions to an anonymized, resource-loaded, large government infrastructure project schedule. We applied triangular probability distributions representing medium uncertainty on all task durations as recommended by the Cost Risk and Uncertainty Analysis Handbook [6]. By viewing the resource utilization heat map, we made several discoveries about the schedule (noted in Figure 2). Based on these discoveries, the project manager may want to reduce resource utilization uncertainty in 2009-Q4 and 2010-Q1, add resources to prevent the highly confident overutilization at the beginning of the schedule, allocate free resources to different projects, and add necessary schedule dependencies.

3 Future Work and Conclusions

Our collaborators (schedule analysts and project managers) enthusiastically received this new visualization, and commented that they are "not aware of any other tool that can perform this kind of analysis." This visualization is deployed in the commercial version of Polaris that is used by over 30 government and industry projects.

Future work can add more sorting criteria for the heat map that consider distance between the resource utilization distributions, such as Kolmogorov-Smirnov distance or Kullback-Leibler divergence. Additionally, resource utilization heat maps could be tied to task-oriented views such as Gantt charts or critical path trees to show when tasks may suffer delays due to insufficient resources.

References

[1] J. Belien and E. Demeulemeester. Building cyclic master surgery schedules with leveled resulting bed occupancy. *European J. of Operational Research*, 176:1185–1204, 2007.

[2] R. Gove, N. Gramsky, R. Kirby, E. Sefer, A. Sopan, C. Dunne, B. Shneiderman, and M. Taieb-Maimon. Netvisia: Heat map & matrix visualization of dynamic social network statistics. In *Proc. of IEEE Conf. on Social Computing*, pages 19–26, 2011.

[3] C. Holzhuter, H.-J. Schulz, and H. Schumann. Enriched heatmaps for visualizing uncertainty in microarray data. In *Eurographics Workshop on Visual Computing for Biomedicine*, Leipzig, Germany, 2010.

[4] K. Molokken and M. Jorgensen. A review of software surveys on software effort estimation. In *Proc. of the Int. Sym. on Empirical Software Engineering*, pages 223–230, 2003.

[5] A. Sopan, M. Freier, M. Taieb-Maimon, C. Plaisant, J. Golbeck, and B. Shneiderman. Exploring data distributions: Visual design and evaluation. *Int. J. of Human-Computer Interaction*, 29(2):77–95, 2013.

[6] United States Air Force. Cost risk and uncertainty analysis handbook. Technical report, Tecolote Research, Inc., 2007.