

Visualizations for User-supported State Space Exploration of Goal Models: Supplemental Material

Yesugen Baatartogtokh, Irene Foster, Alicia M. Grubb
Smith College

This is the supplemental material for the research paper entitled, "Visualizations for User-supported State Space Exploration of Goal Models". This paper is published in the Proceedings of the 31st IEEE International Requirements Engineering Conference (RE). This paper presents a technique for valuation-based filtering and coloring to assist users in understanding a solution space and selecting custom states from it. This supplement contains the data from our initial evaluation and associated models.

Purpose of this Supplement

This supplement can be used review the source data for the tables in the paper. We also provide the models and simulation files for reference.

Preconditions

In order to run the steps below to replicate the example in the paper, users must first get BloomingLeaf working on their local machine. We recommend Release 2.5 of BloomingLeaf, though we used commit [65c61da](#) to generate the elements in this supplement.

Evaluation Visualization Overlay

- Evaluation Visualization Overlay (EVO) is an extension to BloomingLeaf which colors the initial evidence pairs of each intention in a goal model and, after running analysis, overlays the proportion of time points that each intention holds each evaluation. You can turn on EVO in any window by moving the slider next to `EVO` in the navigation bar.
 - Modeling mode has EVO state mode which colors intentions based on their initial evidence pairs.
 - Once `simulate path` is run in Analysis mode the following EVO options are available:
 - EVO % mode which shows an overview of the evidence pair composition of the simulated path for each intention ordered left to right from most to least fulfilled.
 - EVO time mode shows an overview of the evidence pair composition for each intention ordered by its values over time.
 - EVO state mode colors intentions based on their evidence pair at that time point. In state mode you can scroll through the time points on the bottom of the screen to see the coloring of the intentions at each time point.
 - In Next States there is EVO % mode which shows an overview of the evidence pair composition of the state space for each intention ordered by evidence pair and EVO State mode which colors intentions based on their evidence pair value.

Next States

- After simulating a path for the model, you can run Next States, which is an extension to the current analysis algorithm that enables users to explore any time point in a single path solution.
- To explore Next States, move the time point slider at the bottom of the page to any time point from which you wish to explore the solution space and click `Explore States`. This will open a new window. You cannot Explore States from the last time point.
 - The time Next States takes will depend on the complexity of the model. When the scenario has not been fully explored by the user or set up correctly, some models will have too many states to find a solution in a human-in-the-loop timeframe and create a stack depth error.
 - The total number of states generated by Next States is listed at the top of the left sidebar, and the user is able to click through the states using the arrows or use the search function to go to a specific page.
 - The filters on the left sidebar are applied to the entirety of the model, whereas the Intention Filters on the right side of the window can be used to apply filters on individual intentions after you have selected one. You can add multiple filters to different intentions, and they may be added and removed.
 - You are able to `Save & Close` the Next States window after selecting a desired solution, which generates the remainder of the path, click `Explore Next States` to generate all possible solutions for the next time point, given the updated selection, or simply `Close` the window.

Reproducing the Example in the Paper

- To reproduce the examples shown in the paper, the .json files of the analysis have been provided in the “Debt-Analysis-Files” folder. The .json files include the model, path request and response, state request and response, and update request and response. The conflict prevention level, which can be assigned in the drop-down menu on the sidebar in analysis mode, was set to `none`. Please note that there is one randomized time point in the simulation, so if run independently the random time point may change and the time point used in the example could be off by +/- one.

Models for Evaluation Included in the Supplemental Materials

- The following five models are the models we used for our Intention Filtering evaluation. The supplemental materials file contains their .json file which can be loaded into BloomingLeaf to replicate our results for the reduction of states achieved when applying Intention Filter. For our evaluation we used the default conflict prevention level of `strong` which can be assigned in the drop-down menu on the sidebar in analysis mode. Note that the number of possible next states generated may differ for different simulation results.
 - For all models, the filters lab members applied as well as their evidence pairs are listed in the “raw-data.xlsx” file, under the `Filter-Data-Raw` tab.
 - 2022 Biden-Harris Administration Student Debt Relief Plan (Debt Model)
 - The developers for the 2022 Biden-Harris Administration Student Debt Relief Plan are deciding whether they should manually or automatically review borrowers’ IRS Form 1040 to verify that they are within the income limit. If they choose to automatically review, they must decide whether to build the algorithm in-house or to contract an existing 1040 reader. If they choose to conduct a manual review, they must first choose whether to hire and train new staff or limit themselves to their existing staff and then choose whether to create a fax or online application form.
 - To replicate our results, simulate single path with one relative time point and open Next States from the third time point.
 - The filters we applied, in order, using Intention Filter for the worst case were:
 - `Get contract with vendor` as fully denied
 - `Accuracy of review` as partially satisfied
 - `Prepare for future relief` as partially denied
 - The filters we applied, in order, using Intention Filter for the best case were:
 - `Create job opportunities` as partially denied
 - `Build review algorithm` as fully denied
 - `Get contract with vendor` as fully satisfied
 - Course Model
 - A student is registering for classes for next semester. They have space in their schedule for one elective and they are trying to decide if they should take a fun elective which they have an interest in or an useful elective that teaches them a real world skill.
 - To replicate our results, simulate single path with five relative time points and open Next States from the first time point.
 - The filters we applied in order using Intention Filter for the worst case were:
 - `Research class offerings` as fully satisfied
 - `Meet with advisee` as fully denied
 - After two filters were applied, all of the intentions had a single evidence pair or were conflicted for all of the states, so no more filters could be applied.
 - The filters we applied in order using Intention Filter for the best case were:
 - `Learn` as partially satisfied
 - `Meet with advisee` as fully satisfied
 - `Research course offerings` as fully satisfied

- Ready4Work Model
 - An employee is modeling getting ready for work, they have to plan their morning to get to work on time and need to decide if they should make their lunch or buy their lunch and whether they should take the bus or drive to work.
 - To replicate our results, simulate single path with four relative time points and open Next States from the third time point.
 - The filters we applied in order using Intention Filter for the worst case were:
 - 'Eat breakfast' as fully satisfied
 - 'Time' as fully satisfied
 - 'Be environmentally friendly' as partially denied
 - The filters we applied in order using Intention Filter for the best case were:
 - 'Get to work on time' as fully denied
 - 'Buy lunch' as fully denied
 - 'Be environmentally friendly' as partially satisfied
- Grad Model
 - A student beginning their undergraduate career is planning out their goal of getting accepted into graduate school. To do so, they need to both complete their undergraduate degree and have a successful application, with obtaining good grades contributing to their admission. This goal model helps the student identify the tasks and goals needed to be admitted, while taking into account their recommenders.
 - To replicate our results, simulate single path with six relative time points and open Next States from the fourth time point.
 - The filters we applied in order using Intention Filter for the worst case were:
 - 'Submit recommendation letters' as partially satisfied
 - 'Have successful application' as partially satisfied
 - 'Obtain good grades' as fully satisfied
 - The filters we applied in order using Intention Filter for the best case were:
 - 'Have successful application' as partially satisfied
 - 'Finish purpose of statement' as fully satisfied
 - 'Gain experience' as fully satisfied
- WME Model
 - A city wants to manage its waste. To make this happen, they are deciding whether they should use a new dump or continue to use the current dump. If the city chooses to use a new dump, they must decide if the dump should be large or small, given their soft goal of reducing operating costs. Taking into account the city's citizens and their waste separation habits, the city also has the task of processing green waste.
 - To replicate our results, simulate single path with seven relative time points and open Next States from the second time point.
 - The filters we applied in order using Intention Filter for the worst case were:
 - 'GW education program' as fully denied
 - 'Quality of waste separation' as fully denied
 - 'Build large dump' as fully satisfied
 - The filters we applied in order using Intention Filter for the best case were:
 - 'Environmental concern' as fully satisfied
 - 'Process green waste' as fully satisfied
 - After two filters were applied, all of the intentions had a single evidence pair or were conflicted for all of the states, so no more filters could be applied.
- The Scheduler model was used for our EVO evaluation with lab members. For our evaluation we used the default conflict prevention level of 'strong' which can be assigned in the drop-down menu on the sidebar in analysis mode. Note that the number of possible next states generated may differ for different simulation results. Lab members were asked to go through the Next State window, paying attention to one hundred states consecutively while keeping track of zero, one, two close, two distant, and three intentions.

Data Included in Supplemental Material

- The data from our evaluation can be found in the raw-data.xlsx
 - `EVO-State-Data-Raw` tab contains the results of our trials to measure how long it takes to look at states with and without EVO for each of the three groups: Author, EVO First, and Base First. The EVO First and Base First groups had 3 people each. Individuals looked at finding specific evidence pairs for zero, one, two, or three intentions, including two intentions that were close and distant from one another, as well as three intentions equally dispersed throughout the model, to observe possible cognitive processing delays. Author data was originally collected as states per second, as we measured how long it took to go through the entirety of a solution space. We then divided this by the number of states in the solution space to get states per second. However, for the EVO First and Base First groups, we measured how long it took for individuals to go through 100 states. This was then converted to states per second.
 - `EVO-State-Pivot-Table` tab lists the averaged states per second for each group and for each number of intentions tracked, calculated from the individual observations in the `EVO-State-Data-Raw` tab. The Grand Total row lists the average states per second for number of intentions tracked, regardless of group or whether EVO was used. The Grand Total column lists the average states per second for each group, regardless of number of intentions tracked. The bolded table values correspond to the averaged states per second for the separate groups and number of intentions tracked, regardless of whether EVO was or was not used.
 - `EVO-State-Max-Min-Pivot-Table` shows the maximum and minimum number of states per second for each of the groups, whether EVO was first or not, as well as the number of intentions tracked. The Grand Total row lists the average max or min states per second for number of intentions tracked, regardless of group or whether EVO was used. The Grand Total column lists the average max or min states per second for each group, regardless of number of intentions tracked. The bolded table values correspond to the max or min averaged states per second for the separate groups and number of intentions tracked, regardless of whether EVO was or was not used.
 - `Filter-Data-Raw` tab contains the results of our trials with using Intention Filter to reduce the number of states. We performed this trial for five models. For the Author filtering, we performed a best-case scenario where we filtered out as many solutions as we could and a worst-case scenario where we filtered out as few solutions as possible, based on visual examination of the EVO % mode. For each scenario we applied three filters, recording the number of states each filter reduced the number of solutions by, as well as the number of starting solutions. People 1-5 were members of our lab who were asked to apply three filters to the same five models, based on their intuition and what they wished to achieve after examining the models. It should be noted that a large percentage of solutions contained conflicted values, which you cannot filter by. Thus, our filtering was over the five initial evidence pairs. This table includes both the number of states remaining and percentage reduction for each observation, as well as the choice of intention selected, and filter applied.
 - `Filter-Data-Pivot-Tables` tab contains the tables calculating the average of percentage reductions for the author and lab members' filtering. In the first table, the Grand Total column values correspond to the average percentage reduction across all five models. The Grand Total row corresponds to the average percentage reduction per number of filters applied. The second table is similar, except it includes the Author measurements of the Best and Worst case and the grand averages of that.