A Tool for Visualising Cell Model Results
MInf Project Phase 2

James Hulme
s0901522

January 13, 2014

Contents

1 Introduction

Biologists often use computer models to help guide their research as they are much cheaper than doing the experiments for real. There are a number of tools available for their use. These tools typically require a certain level of numerical confidence to create and interpret the results from. Not all biologists have this numerical confidence. For some writing and interpreting the model is a challenge, this can make them less effective in their research. It is therefore necessary for these tools to help them relate the data to their field by incorporating domain knowledge.

One such tool that can be used for modelling is Bio-PEPA, an extension of the PEPA process algebra. Bio-PEPA is currently implemented as a plugin for Eclipse IDE. Bio-PEPA visualises data as time-series graphs. There is one team in particular who use Bio-PEPA and are not comfortable with analysing graphs. This team has provided the focus for this project.

The purpose of this project was to extend Bio-PEPA's visualisation capabilities to allow the previously mentioned team, and other similar users of Bio-PEPA to more effectively analyse their results.

A significant problem with Bio-PEPA's visualisation capability is that it is difficult to represent spatial change on a time-series graph. In Bio-PEPA you can have a species at different locations in the cell, for example, next to the nucleus, next to the cell membrane and throughout the cytoplasm. The movement of this species through the cell can by modelled by seeing the population of it in each location over time. This is difficult to visualise on a time series plot as three lines is too abstract. It requires the use of biological metaphor to be easily interpreted. In this case a visualisation of the cell that can show how the species moves through the cell.

Over the course of the project the scope has been expanded. The original objective was to assess which forms of visual representation are most helpful and informative to laboratory science. At the end of the first project phase the object changed to be to develop a tool to visualise the results of dynamic, time-series models of intra-cellular behaviour based on biochemical reactions. This objective was focused on visualisation to aid those researchers who are not numerically confident. The second phase of the project has added to this objective to also aid interpretation and collaboration. This change in objective is to make the tool better for all users.

1.1 Where Does This Tool Fit In?

In the first stage of the project a review was performed of the features of a number of modelling and visualisation tools. This review included specialised software aimed at biologists and general software for anybody doing data visualisation.

The software that was reviewed were: Bio-PEPA, Uppaal, V-Cell, Cell-O, Copasi, Cell Designer and WEAVE. Bio-PEPA, V-Cell, Cell-O, Copasi and Cell Designer have been written for biological modelling. Uppaal is modelling software for general use, and WEAVE is a general data visualisation tool.

All offered some level of visualisation, some simply graphs, others more complex visualisations. Bio-PEPA offered only line graphs. Uppaal had visualisations that highlighted where in a finite state machine view of the model the current state is. V-Cell had visualisation of the model in a hierarchical set of circles, it was also display spatial elements of the results data by displaying a heat model view of the cell. Cell-O was aimed more at multi-cell models and was able to show them moving and splitting, it also had visualisation of the model as a finite state machine.

Copasi only graphs although the user had more control over the display of the plots. WEAVE had the largest visualisation capacity being able to display a variety of standard graph times along with more interesting ones, such as geographical maps, but it did not appear to have anything specialised for biological models. WEAVE is also the tool that gave the user the most control over the visualisation.

The existing biological modelling software seems to be focused more on the ease of modelling. The visualisation features on offer are typically quite basic. They also lack the more innovative features that can be found in the newer general data visualisation software.

An important area that all the reviewed software were lacking in is collaboration. In recent years cloud software has taken off and made real time collaboration a possibility. The current work flow using the existing software is do your analysis, save it, email it to a colleague with additional files and notes, have them open it and try and work out what is happening. It is a disjointed conversation. None of the tools surveyed offered a more efficient approach.

This tool has a space in the existing landscape in two ways. The first as offering visualisation capabilities similar to those in the other biological modelling software to Bio-PEPA. The second is by implementing modern collaboration features, making it a unique tool.

1.2 Previous Work

Early on in the first stage of the project it was decided to separate this project from the Eclipse plugin. It was felt that Eclipse is not the right tool to do data visualisation in.

The initial development stages were focused on getting the new tool from zero functionality to matching the visualisation features of the Eclipse plug-in. This involved writing an early version of the U.I., parsing the Bio-PEPA results data, and plotting it using matplotlib.

After matching the existing functionality it was time to add new functionality. The first new feature was intensity plotting where the colour of the line increases in intensity/opaqueness as the population of the species increases. Then visualisation of the model was added. It used a system of nesting circles and rings to build a heirarchical view of the cell from its model components. Finally the user was given control of the plot, allowing the toggle whether it is shown or not, how it is plotted, what colour it is and the thickness of the line.

Over the course of the first stage of work a number of evaluations were carried out with potential and actual users of Bio-PEPA. The findings from these evaluations were used to improve the tool.

1.3 Results

Break down by project stage? Aniticipating this to be 1.5 - 2 pages

2 Goals

Whilst researching the problem and the existing software, and during the first stage of development a number of goals were identified.

- Improve existing capabilities

- Visualise the model more intuitively
- Visualise closer to the cell level
- Animation of the data
- Investigating which combinations work best
- Add the ability to annotate the visualisations
- Allow the ability to save and load the program state
- Provide a full session history to the user
- Data Mining
- Making meta data accessible

During the first half of the second phase of development new goals were identified.

- Google docs style collaboration.
- Searchable database of time series plots.
- Manipulation and Exportation of data.

2.1 Previous Work

Could move here from intro, could talk a bit more about how they could be improved.

2.2 New Work

For each of the below, why, how inc conceptual problemss, impact, how it could be improved

2.2.1 Annotation

2.2.2 Data Mining

2.2.3 Search

2.2.4 Collaboration

2.2.5 Usability

2.2.6 Animation

2.2.7 Data Manipulation and Export

3 Evaluation

3.1 First Evaluation

The first evaluation in the second phase of the project occurred just before the halfway mark. The group was made up of two people. One who had taken part in the first evaluation meeting and one person who had no knowledge of the project.

3.1.1 User Group

As I did not have a domain expert available I was not able to do insight based evaluation. I took a more traditional approach. Before the evaluation I prepared a typical scenario that a user might encounter. The task was to open a file, annotate it and play the animation, and attach some supporting documentation. The task was prepared at two levels of breakdown. One was a paragraph of text at quite a high level. The second level of breakdown was a step by step instruction of each action to perform. I then observed them as they attempted the task and offered assistance when required. Afterwards I gave them a questionnaire to fill in about their experience, afterwords we went through and discussed their answers and any further thoughts that they had.

The task was prepared at two levels to try and gauge how easy the program is too use. The users were presented the textual description and if they had struggled they would have been given the step by step instructions instead. The users were able to complete the task from the textual description alone. This is a good sign that the new tool is usable.

Some issues were encountered:

- Being unfamiliar with MacOS – Both users were unable to locate the menu bar as it is not attached to the program as in Windows. Future evaluations will use Windows.
- When annotating they were unclear as to what was going to happen when annotating. For example when annotating the graph with an arrow the user was just left to click twice, with no indication of what would happen. This has now been fixed, different cursors are used to give feedback to the user that they should click, and rather than just relying on two clicks with no information as to where the arrow is going to point, after the first click

(which places the tail of the arrow) an arrow will be drawn that follows the cursor until the second click placing the annotation.

- Lack of ability to edit, move, or delete annotations – Once an annotation was placed it was there for good. The ability to edit annotations was always planned, it just had not been implemented in time. But the amount of frustration it gave the users was very high. It was a principle in all three of the design lists that a user should be able to fix mistakes. Since the evaluation editing and deleting of annotations have been implemented. This means any mistakes can be corrected.
- Initially they were confused by what all the buttons on the matplotlib toolbar did. After discovering the tooltips and seeing what effect the buttons had they were comfortable with them. Also all the matplotlib built in buttons on the toolbar are undoable from the toolbar. Annotations are editable and deletable and also covered by the undo functionality implemented across the whole program. Being able to recover from their actions on the toolbar means no hindrance to discovery and so needs no further action.
- The users were confused by some of the terminology. In particular "save graph" and "save model". These items in the menu have grouped more carefully to help the user distinguish them. A related issue was worrying that "save graph" was going to override the results file. To rectify this the menu items that create new files have been renamed "export ...".
- When creating a session they stuggled, they did not know what the title was referring to. And when trying to add files, rather than use the add files button in the dialog, they tried to use the file menu. Having two routes into doing the visualisation seemed to be confusing them. Now the file menu open file has been removed. To create a visualisation the user has to go through the new session wizard.
- When placing species in the cell one of the users did not understand what they were being asked to do. One of the users did understand. To fix this user input has been removed from the equation. This has required the model file to also be chosen, but then species locations are parsed automatically.
- They liked the animation feature and thought it would be very useful (One of the users did their phd in transport and expressed a desire to have had this feature during the phd). They did feel that it wouldn't be useful directly for papers, but that it could be used to inform the content of papers.
- One of the users asked if they had a map of the cell. When presented with the model visualisation they thought that it did look nice, but were unsure of its usefulness – HAS IT BEEN TAKEN OUT.