

Project Proposal - APT467

Andrew P Talbot 1434667

June 22, 2018

1 Project Title

A Potential Project title could be "A KNIME Extension for the Analysis of Mass Spectrometry Imaging Data, Exploring Topological Techniques to Improve Clustering Methods on Mass Spectrometry Data."

2 Aim

I will create an extension of the open source data analytics/workflow platform KNIME which aims to implement accessible Mass Spectrometry Imaging (MSI) analysis in an extensible manner. Then I will explore topological approaches to clustering, in an attempt to improve algorithms that are currently used to cluster MSI data.

3 Plan

The project is split into the two following pieces:

- The development of a KNIME extension to allow easy, accessible analysis tools to be used for MSI datasets.
- Exploring topological approaches to the clustering of MSI datasets.

Here, I will explain how the project can be divided into small pieces and a rough time line to expect them to be finished is provided in brackets at the end of the bullet point. Note that, these tasks and the time line will likely evolve as the project progresses.

3.1 KNIME Extension

The bulk of this part of the project aims to be completed before moving on. The majority of which should be completed by 20 July 2018. The code for this will be written in Java as that is what KNIME is written in, and is the language that I am most proficient with.

The tasks that need to be completed for this part of the project are:

- Software process, requirements, architecture and documentation for the project, this in particular should be done in an agile fashion, so this will not be finalized until perhaps the end of this part of the project. However, a solid beginning should be made early. (26/06/2018)
- A parser node for imzML data is to be integrated with KNIME for loading MSI data into a KNIME accepted format. (29/06/2018)
- Nodes are to be created for methods that can be used to pre-process/analyze the MSI data. (15/06/2018)
- If there is time available, some time should be spent creating KNIME nodes that allow MSI data to be visualized. (22/07/2018)

3.2 Persistent Clustering

This is the majority of the project. This is to be started in early July, and completed by the end of August. Some self teaching of clustering algorithms and topology should be done at the start of this part of the project, and will perhaps be ongoing for the duration of the project.

The tasks that need to be completed for this part of the project are :

- A Literature review and a review of topology basics needs to be completed before the bulk of this work is done. (31/07/2018)
- Convert the algorithm into a KNIME node, so that it can be used repeatedly, and added to the software. (10/08/2018)
- Compare conventional clustering algorithms to the topological algorithms explored in this project, and test the hypothesis that it will have a lower error, over the same datasets. (19/08/2018)
- Discuss future areas of research/methods to improve the topological clustering algorithms. Perhaps a great deal of time could be spent writing the report in this week also. (31/08/2018)

3.3 Report Writing

I aim to begin this part of the project early, but the bulk of it can begin in the middle of July (perhaps on 16th July 2018). This can be done alongside the other parts of the project. I aim to have finished the report writing by the end of August. So that it can be finalized in the last few days of the project.

3.4 Proof Reading and Final Work

I aim to have the report proof read and have a completed final draft, with the software finished and deployable in an extensible fashion (so that in the future it can be easily improved/gain functionality) by 7th September 2018.