

EEE3094 Technical Report

Semester 1 Report

Sahas Talasila 230057896

CONTENTS

1	Introduction	2
2	Literature Review	2
3	Theoretical Background	2
3.1	Graphs and Time Series Background	3
3.2	Visibility Graph Background	4
3.3	Visibility Graph Limitations Background	4
3.4	Min-Max Heap Background	4
3.5	MOVG Background	4
3.6	ML Pipeline and Preprocessing Background	4
4	Results	4
5	Conclusion	4
6	References	5
7	Appendices	5
8	Detailed Write Up Plan	6

Abstract

This is a placeholder for the abstract of the report.

INTRODUCTION

The Introduction should answer three basic questions:

What is the project about? This section should be a precise description of the subject matter of the report. It should also include the way it relates to other work in the School, if this is appropriate.

Why is it being undertaken? The reasons for undertaking the project are discussed in this section, outlining the objectives and importance of the work. Since it is important not to create a false impression of the outcome of the project a brief statement is made regarding whether or not the aims of the project were achieved, and, if not, why not.

How is the subject matter of the project described? – This should be a ‘Road-Map’ of the dissertation. This is a ‘Road-Map’ of the dissertation and comprises a short summary of the contents of the subsequent sections. It informs the reader where, for example, the methods used, results obtained, etc. are to be found.

LITERATURE REVIEW

The Literature Review defines the current state of research in your area and places your work in context with other work in your project area. It also acts as the foundation for the comparison of your results with other relevant work in the ‘Discussion’ chapter in your Individual Project.

The review is a summary of relevant articles (that is, material relevant to the background to your project, overview of the subject, discussion of any essential theories, etc.) published in technical journals, conference proceedings, books, websites, etc. - avoid using very general references. The majority of references should be books or journal papers; web sites should appear only occasionally in the list.

Most journals have ‘Special Issues’ on given topic areas. These are good sources of review material and contain large numbers of references.

However, do not cite a reference unless you have read it. Furthermore, a Literature Review which only cites Stage I and II lecture notes and text books is not considered to be adequate.

It is imperative to attribute any information cited in the report to the source from which it was obtained - failure to reference material properly could result in plagiarism.

The list of references should be provided in the same format as in IEEE Transactions or IET Proceedings.

THEORETICAL BACKGROUND

This section should provide the theoretical background to the project. It should include any relevant equations, derivations, and explanations of concepts that are essential for understanding the project.

3.1 Graphs and Time Series Background

Define and ask:

1. What is a graph?
2. What is a time series?
3. Where are time series used?
4. What is time series analysis?
5. What are the common methods used for time series analysis?
6. How can graphs be used for time series analysis?
7. Why can graphs be used for time series analysis?
8. What is a visibility graph?
9. How can visibility graphs be used for time series analysis?
10. Why can visibility graphs be used for time series analysis?
11. What are the common methods used for visibility graph analysis?
12. What are the limitations of visibility graphs for time series analysis?
13. What is a Min-Max Heap?
14. How does a Min-Max Heap work?
15. What are the common operations of a Min-Max Heap?
16. What are the time complexities of the common operations of a Min-Max Heap?
17. What are the limitations of a Min-Max Heap?
18. What is the MOVG?
19. How does the MOVG work?
20. How is the MOVG different from a standard visibility graph?
21. What are the advantages of the MOVG over a standard visibility graph?
22. What does the MOVG take from the Min-Max Heap?
23. What are the advantages of the MOVG over a Min-Max Heap?
24. How does the MOVG fit into a machine learning pipeline?
25. Why is preprocessing important for machine learning?
26. What are the advantages of using the MOVG for preprocessing in a machine learning pipeline?
27. What are the possible applications of the MOVG?

3.2 Visibility Graph Background

3.3 Visibility Graph Limitations Background

3.4 Min-Max Heap Background

3.5 MOVG Background

3.6 ML Pipeline and Preprocessing Background

Diagrams to show what the ML model pipeline is and where the MOVG fits in, as well as explaining why preprocessing is important.

RESULTS

Results should be presented in a clear and succinct way so that they can be easily interpreted and also supported by appropriate text which highlights the significant aspects of the results.

If a design project has been undertaken the results will comprise measurements, observations and simulations which have been carried out in order to assess how closely the design correlates to its specification in terms of its functionality, performance, area, power dissipation, etc.

In this section the results of your investigation or design are analysed and compared, if appropriate, to related work in the field. Again, it is also useful to reiterate the aims of the project so that your discussion can be put in context. What ensues thereafter depends upon the project. For example, it may include a comparison of:

- Measured results with simulation data, or results outlined in related technical publications.
- Results from different proposed solutions, different implementations of the same solution – advantages / disadvantages of each.
- Design performance against its specification.

The discussion may also include a Retrospective Section where, considering the knowledge and experience you gained from the project, you may give a brief account of how you would have tackled the project differently.

Suggestions for extending the work of the project can also be outlined in this section.

CONCLUSION

The conclusions should be very succinct and depending on the project may contain:

- Statements on the deductions obtained from the results.
- Comments on the extent to which the original aims of the project were fulfilled.
- Novelty of the solutions proposed - merits and limitations.
- Statement of potential applications.

REFERENCES

The list of references follows immediately after the conclusions. Each reference should be cited somewhere in your dissertation, and there should be no citations that refer to references that cannot be found in the list. Use IEEE standard for referencing.

APPENDICES

should be used for material that is relevant to the project but disrupts the flow of the main text, e.g. detailed derivations, additional graphs, tables of results, program listings, etc.

Need to add API design documentation for the MOVG code

Need to add links for github repo and also pyPI package.

Need to add all of the links and tables for planning this project.

Need to test IEEE format for writing the report

Need to add flowcharts for algorithm design.

Output examples of the MOVG code, including examples to show how the MOVG deals with edge cases, like missing values and NaN values. We can also see how it deals with small and large datasets, without a GUI as well.

Need to include:

- Flat trends,
- Increasing trends (steady and sharp increases),
- Decreasing trends (steady and sharp decreases),
- Staircase trends, up and down,
- Plateaus,
- Missing values,
- NaN values,
- Small datasets,
- Large datasets,
- Seasonal trends,
- Periodic trends,
- Random noise.
- Applying AR or MA methods to the time series before applying the MOVG.
- Applying smoothing methods to the time series before applying the MOVG.
- Applying differencing methods to the time series before applying the MOVG.
- Applying detrending methods to the time series before applying the MOVG.

- Need to use two similar time series, but one with a slight variation, to show how the MOVG can pick up on small changes.
- How clustering can be retained in the MOVG output.
- How outliers can be retained in the MOVG output.

DETAILED WRITE UP PLAN

Get the IEEE format working for the report. IEEEtran seems to be the best package for this.

1. Abstract
2. Introduction
3. Literature Review
4. Theoretical Background
 - Explain what a graph is mathematically, using the correct mathematical notation + source.
 - Explain what a time series is and how it can be used in a financial setting.
 - What is a visibility graph and where can it be used (applications plus original paper as a reference).
 - Explain how a visibility graph can be used for time series applications and analysis.
 - Explain why we can use them as well.
 - (a) Traversal/Partition is not possible due to the amount of edges and the nodes are connected.
 - (b) No clear traversal methods for a standard visibility graph.
 - (c) Cannot see the trends clearly for visibility graph.
 - (d) Compare with MOVG as well.
 - Need to explain what the Min-Max Heap is and how it works, including pseudocode and time complexity analysis.
 - Explain the operation of the Min-Max Heap (used to create a data structure from an array).
 - Min-Max heap capabilities and how it can be improved and downsides (de/-composition isn't actually possible).
 - Need to explain signatures and how they work.
 - Explain what the outputs look like and how it can be fed into a model
 - Start to explain how the MOVG works:
 - (a) Talk about preprocessing capabilities and how it fits into the process of designing a machine learning model.

- (b) What the outputs are and how it can be fed into a model.
- What problem does it aim to solve?
 - What are the possible applications - (Ideally financial applications)?

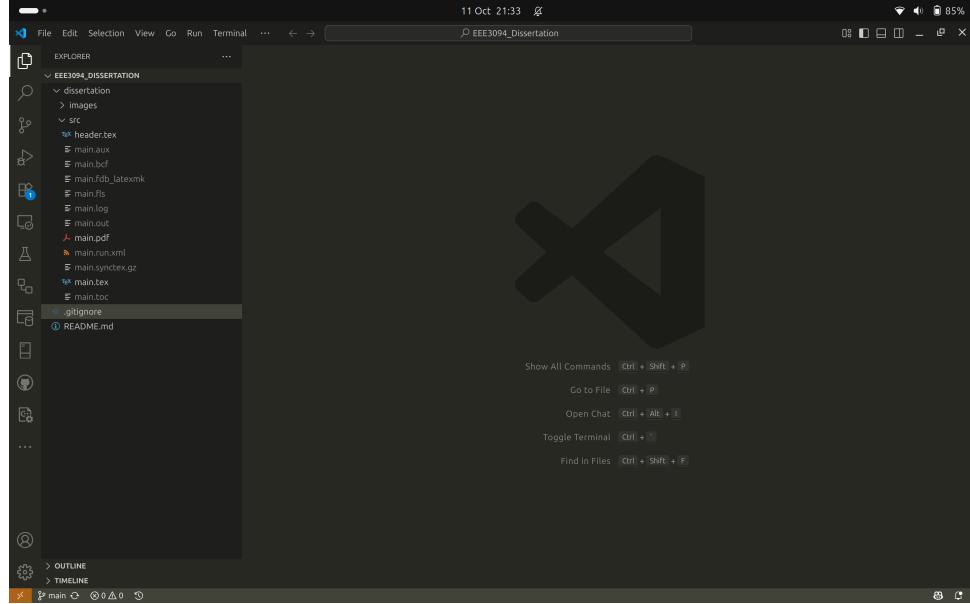


Figure 1: MOVG Diagram