COMPUTING LABORATORY SEMINAR June 18th 2009

Writing an MSc Dissertation

Dr Maggie Charles, Language Centre maggie.charles@lang.ox.ac.uk

1: PARTS OF THE DISSERTATION

Title page

Abstract

A short summary of the whole dissertation, about 200 words.

Acknowledgements (optional)

Thank your supervisor first, followed by any other academic, technical or funding support. Then you can thank your friends and relations.

Table of contents

This should list ALL parts of the dissertation, including References and Appendices

Chapters 1, 2 etc.

Areas to cover: Introduction, Explanation of the Problem or Context, Description of the Method, Account of the Work (several chapters, the main body of the dissertation), Conclusions.

These are the areas given in the Handbook, but you do not have to use these exact words in the chapter titles. Titles should be informative for our own content. You may also find that you can combine certain areas into a single chapter. Many dissertations have around 6-7 chapters, but your choice of organisation should follow from the nature of your own material. The word limit is 30,000, but many dissertations are less than this. 15,000 - 20,000 words or around 50 pages is reasonable. You will probably need about 1,000 words for the Introduction and the same, or less, for the Conclusion. The rest of the chapters will vary, depending on the number you include, but many will probably be around 2-3,000 words. Again the length depends on the content.

References

The list should include ALL the sources that you refer to (print and electronic), giving author, title and publication details. Formatting should be consistent.

Appendices

These give supporting information for your dissertation. They provide extra, non-essential information.

2: INTRODUCTION

- 'Should introduce the subject of the dissertation and explains the structure of the text to the reader'

The introduction is often one of the hardest parts to write. It may be easier to write it last, as you will then have a clear overview of your whole dissertation. It gives the reader some general background to explain and justify the subject of the dissertation. It establishes that there is a problem or need, indicates how the problem is solved and gives an outline of the dissertation structure so that the reader can follow it more easily. There are often three stages (moves) starting with more general information, then establishing the problem and finally narrowing down to specific statements about your own work. These moves may occur more than once.

Introduction

Move 1 Background

Possible Steps
1. Why the area is important
2. Giving background information
3. Reviewing previous research

Move 2 Indicating a Problem or Need

Move 3 Presenting the Research

Possible Steps
1. Purposes, aims, objectives
2. Work carried out
3. Justification or importance of the research
4. Outline of the structure of the dissertation

TASK 1: Investigating an Introduction

- Extract A is part of an introduction to a dissertation.
- Identify the 3 moves and the steps that make them up. The moves may occur more than once.
- Underline the language used to signal these moves and steps.

Extract A Introduction: Modelling of Financial Markets with Intelligent Agents

1.1 Motivation

Financial markets are often viewed as an example of a *complex adaptive system*… In general, a financial market consists of many agents who conduct their activities by a set of rules… From the observer's perspective, the behaviour of every single agent is seemingly random and unpredictable. It is this unpredictability that gives rise to most of the challenges when we attempt to model the whole system…

Back in the early 1990's, the success of the Santa Fe Artificial Stock Market (SF-ASM) [3][5] sparked a significant research interest in simulated financial markets (using ABM). Many new models have been proposed since then [4]. However, the intelligent tools used in most of these works still disregarded current state-of-the-art AI techniques [4]. One reason is that many researchers working in this area do not have a computer science or engineering background. It is this increasing demand for more researchers from outside the economics discipline and the need of more sophisticated intelligent tools that motivated this project.

1.2 Objectives

In this project, we attempt to model an asset market following agent-based modelling approach, where agents are constructed using several artificial intelligence tools: artificial neural network (ANN), neuro-fuzzy system (NFS), and genetic algorithms (GA). Artificial neural networks have also been used by LeBaron [11] and Beltratti et al. [12] in their models but are rejected by other researchers because of the *black-box* nature. We try to remedy this problem by introducing an extra layer to control and monitor the use of network input signals.

The objectives of this project are as follows:
First, we aim to implement an agent-based simulation framework for a stock market…
Secondly, we aim to construct agents using artificial neural networks, neuro-fuzzy systems, and genetic algorithms. Two agent models will be developed: the neural agent model and the neuro-fuzzy agent model…
Finally, we aim to test the agent models by running a number of experiments. Experimental results will be examined through statistical analysis and be compared with the real stock market.

This thesis consists of six chapters. Chapter 1 presents a brief introduction of this project and the thesis structure. The other chapters are described as follows:
Chapter 2 covers the theoretical background of the intelligent systems involved in modelling the agents in this project. It includes the basic theories of artificial neural networks, neuro-fuzzy systems and genetic algorithms.
Chapter 3 presents the related work, namely the SF-ASM model.
Chapter 4 describes in detail our modelling approach.
Chapter 5 describes the concrete work of this project. It includes a brief description of the simulation framework, out experiments, and statistical analysis results.
Chapter 6 summarizes the project results, contributions, and directions of potential future work.

3: EXPLANATION OF THE PROBLEM

- 'Should explain the problem to be studied or the context in which the work takes place'

Extract B: Machine Learning Techniques for Astrophysical Modelling and Photometric Redshift Estimation of Quasars in Optical Sky Surveys

2.1. The Orbits of the Satellites of Jupiter

…**The problem undertaken here is, in principle, to determine** all of the six orbital elements [7, p. 58] that precisely describe the orbit of a given satellite around Jupiter, **using** only observations available from telescopes on Earth. **Given** the well defined Keplerian theory of celestial mechanics, **this problem is essentially one of function approximation** in six dimensions. **As such, it is similar in form to any other classical modelling problem** in astrophysics, and **demonstrates the applicability of** learning techniques to the approximation of highly parametric models.

More immediately, however, **this solution** of approximating celestial orbits **can be used to study** binary star systems, comets with highly eccentric orbits, and planetary star systems, as well as astrodynamical problems such as optical or lossy radar-based tracking of missiles and satellites…

3: EXPLANATION OF THE PROBLEM (continued)

2.4. The Kepler Problem
The calculation of position in an orbit (in terms of the orbital elements described in 2.3) as a function of time is known as the Kepler problem,[5] as it requires the solution of Kepler's equation [7, pp. 185, 193]:

(2.1) $M = Mepoch + n \cdot \Delta t = E \_ e \cdot sin(E)$

where *n* is the *mean motion*,6 Δ*t*7 is the time elapsed since the current epoch, *E* is the eccentric anomaly [7, p. 183], another measure of progression in orbit, and *Mepoch* and *e* are as above.

4: DESCRIPTION OF THE METHOD
- 'Should introduce the method used to solve the problem or the formal techniques employed'

Extract C Spatial Domain Steganalysis using Statistical Learning Techniques

2.2 Proposed Methods
Once the domain for the proposed steganalytic system has been agreed, the system designer is required to make the following **design choices**:
Firstly, the choice of features, i.e. what statistical properties of the images are to be extracted for the analysis.
Secondly, the choice of the classifier, i.e. the particular machine learning tool that will be trained on the chosen features and subsequently used for predicting whether a given image contains payload or not.
The idea is to produce a system that predicts the class of an image, stego or cover (unaltered) image, with high accuracy…

2.2.3 Techniques for Experiments
In order to ensure fair testing conditions and, as a result, **conduct a sound comparison** of the classifiers, their **performance was analysed using** a set of formal experiments.

Firstly, ten-fold cross-validation **was employed to ensure that** the measure of a model's performance represents the general performance of the classifier on a given data set with some accuracy. To reduce the probability of a test set being by chance more favourable for one of the algorithms in question, **homogeneity** of training and test sets **was imposed** using stratification, i.e. the **examples were randomly distributed** across the data sets. In the resulting sets, **the instances were drawn from** (roughly) the same distribution. Finally, the Student's T-test and the Wilcoxon signed-ranks test **provided statistical evidence** for the conclusion.

5: ACCOUNT OF THE WORK
- 'Should present the work carried out, including any practical results and theoretical insights obtained'

These are the main body chapters of your dissertation and will therefore vary substantially according to the exact type of work you are carrying out. They are likely to be the longest chapters, probably over 2,000 words each. Each chapter is organised into sections and sub-sections, again of varying length.

Extract D: Efficient Query Evaluation for Tuple-Independent Probabilistic Databases

Table of Contents
3 Hierarchical Queries Induced by Functional Dependencies

4 Efficient Evaluation of Hierarchical Queries using OBDDs
4.1 One-Occurrence Normal Form (1OF)
4.2 Query Signatures
4.3 Good Variable Orders

5 Exploiting 1OF and OBDDs: The HQ Operator
5.1 Incremental Confidence Computation on OBDDs
5.2 HQ Semantics
5.3 HQ Optimizations
5.4 HQ Evaluation

5: ACCOUNT OF THE WORK (continued)
6 Implementation of HQ in PostgreSQL
6.1 Integrating HQ into MayBMS
6.2 Support for Tuple-Independent Probabilistic Relations
6.3 New SQL Constructs: conf and tuple conf
6.4 Query Execution

7 Case Study: Probabilistic TPC-H

8 Experiments
8.1 TPC-H Data and Queries
8.2 HQ's Engine vs MystiQ's Engine
8.3 Lazy Plans vs Eager Plans vs MystiQ Plans
8.4 Lazy Plans vs Eager Plans vs Hybrid Plans

- In the following example, notice how the writer points out the important information to the reader and how he/she structures the section logically and signals it explicitly.

4.2 Query Signatures
A query signature, also called variable order type [15], represents a class of variable orders. The formal definition of query signatures is:
Definition 4.4. *A query signature is*
- *a table name X defining a class of variable orders with only one variable from table X*
- *α⁎, where α is a query signature, defining a class of variable orders concatenating at least one variable order with query signature α*
- *αβ defining a class of variable orders that concatenate two variable orders with signatures α and β respectively.*

A table name X can only occur once in a signature.

There are four points worth noticing from the definition above. Firstly, it is possible that two syntactically distinct signatures are semantically identical. **To be exact,** several consecutive stars(*) and a single one have the same effects. **For instance,** X* and (X*)* both define the concatenation of variable orders in the form of X. The outermost star in the second signature does not take effect. **Secondly,** multiple occurrences of the same table are not allowed. **As a result,** the framework in this section is not suitable for queries with self-join. **For example,** q:- R, R is not a PTIME query and fails to admit a good variable order. **Thirdly,** due to the inductive nature of the definition, stars can follow any signature. These may be extremely complicated and even have nested stars. **A good example** is ((A*B*)*(C*D*)*)*, in which elimination of any of the 7 stars semantically alters the signature. **Lastly,** X is a refined signature of X*. It is always safe to replace X with X* because X* requires at least one variable; **however,** X is more informative than X*.

Example 4.5.

6. CONCLUSIONS
- 'Should contain conclusions drawn following the project, comparisons between this and existing work or practice, and suggestions regarding the extension or continuation of the work'

In the 'Conclusions' you summarise your work and then stand back from it and assess it as objectively as you can. This shows the examiners that you are capable of evaluating your own work according to the standards of the field. It is acceptable to mention areas that were not very satisfactory because this shows what you have learnt from carrying out the research.
There are often three moves, starting with a summary, followed by an evaluation, highlighting achievements and limitations and finally future extensions of the work. The Evaluation move may occur as part of the Summary. The Conclusion is the mirror image of the Introduction; it starts with the narrow concerns of the research and widens out to more general statements about further work.

Move 1: Summary
Summarises the most important aspects of the work.
Move 2: Evaluation
Indicates what the writer considers to be the most valuable contribution of the dissertation and its limitations. Provides an assessment of how far the aims of the dissertation have been achieved.
Move 3: Future Work
Gives possible extensions of the dissertation. These suggestions may follow from the limitations mentioned in the Evaluation move.

TASK 2: Investigating the Conclusion

- Extract E is part of the conclusion to a dissertation and shows the first 2 moves. Extract F shows the 'Future Work' move.
- Underline the language used to signal the three moves.
- What do you notice about the tenses and verb forms used?

Extract E The Reachability Analysis of Timed Automata

7. Conclusions and Future Work

The forward reachability analysis is implemented successfully by running a large number of timed automata examples. Simulation shows the program is capable of examining all timed automata states, and the program can always terminate correctly by generating reports which outline what states have been visited and what states have not been visited. In another way, the project also proves that the forward reachability analysis using DBM zones can reduce the infinite DBM states to a finite set of states…

By working on this project, I have had the opportunity to improve my research skills by studying advanced mathematics and theoretical knowledge. My experience in object oriented programming has further developed. I have applied the knowledge learnt in the masters course such as MVC pattern, polymorphism etc to the project work.

To sum up, the project has fulfilled all of its objectives in terms of technical skills developed and soft skills enhanced.

Extract F
Machine Learning Techniques for Astrophysical Modelling and Photometric Redshift Estimation of Quasars in Optical Sky Surveys

Chapter 9 Implications of Research and Future Directions

Part II demonstrated that trivial applications of machine learning techniques can yield results comparable to the most advanced applications of traditional techniques in photometric redshift estimation. As mentioned in earlier chapters, the improvement of these learning techniques - for quasars in particular - will greatly assist research in cosmology [32] and the large-scale structure of the universe [47]. It would be useful to determine the relationship between photometric redshift accuracy and quasar magnitude, as the efficient estimation of faint quasar redshifts would assist in studying quasar evolution [32]. It would also be interesting to investigate in more detail the statistical properties of the z_{phot}-z_{spec} relation and the CZR degeneracy.

Improvements may also be made by calculating the Jacobian matrix for RBFNs as in §6.2.1 and comparing its z_{phot} deviations σ_y with those obtained by ANNz… However… as not all science applications require minimal redshift errors [32], serious consideration should also be given to the unsupervised learning properties of RBFNs insofar as they may be of particular advantage in large sky survey science in the future.

7: ABSTRACT
The abstract provides a summary of the dissertation. About 150-200 words should be enough. It may contain some or all of the following moves, often, but not always, in the following order. Only use one or two sentences for each move. Keep the language and sentence structure simple. The reader should be able to read the abstract and obtain the necessary information quickly and efficiently.

Move 1: Background to the Dissertation
Move 2: Statement of the Problem/Gap in the Research
Move 3: Purpose of the Dissertation
Move 4: Work carried out/Methods
Move 5: Results
Move 6: Conclusions/Implications
Move 7: Contribution of the Dissertation

TASK 3: Investigating the Abstract

- Identify the moves used in abstracts 1 and 2 and underline the language which signals the moves.
- Which abstract do you prefer and why?

Abstract 1: Spatial Domain Steganalysis using Statistical Learning Techniques

Research in the field of steganography and steganalysis is concentrated around finding and benchmarking steganographic schemes and their steganalytic counterparts. A particularly challenging steganographic scheme is LSB matching and, to date, only one system, called the Wavelet Absolute Moment (WAM), exhibits truly sensitive detection of it. This thesis is an investigation into the limitations and potential improvements of the WAM steganalyser.

Firstly, we incorporate two more powerful classifiers, namely support-vector machines and neural networks and perform statistical tests against a selection of data sets. The resulting benchmark conforms sensitive detection, but also reveals issues regarding WAM's stability on certain types of images.

Secondly, we simplify the WAM model by adding more sophisticated features to the original WAM feature set and performing simple feature selection. The result of this work is believed to have made a strong contribution to the field. (139 words)

Abstract 2: *k*-best Parsing Algorithms for a Natural Language Parser

The implementation of four algorithms allowing the C&C natural language parser to output multiple ranked parses is considered. The background to the problem is examined, along with the motivation behind the goal of *k*-best parsing. The parser's current implementation is described, along with the changes necessary to add these new algorithms. Finally, a series of tests is conducted to evaluate the efficiency and potential utility of the *k*-best parsing algorithms. (70 words)

8: AVOIDING PLAGIARISM

Plagiarism is defined as the use of other people's work and the submission of it as though it were one's own work. It may be intentional or unintentional.

In academic writing, it must be clear to the reader at any given point whose 'voice' is speaking. This is called **transparent source use**.

- **The responsibility of transparency in writing** means that the writer must use appropriate signals so that an experienced academic reader can understand the actual relationship between the source text and the new one.
- **The responsibility of transparency in language** means that language which is not signalled as a quotation is understood to be original to the writer. If the content is marked with a citation, the language is understood to be paraphrased, i.e., **substantially and independently reworded**.

The process of signalling the presence of another writer's voice in your text is accomplished through **citation**. Citation includes both **paraphrase with reference** and **quotation**.

A paraphrase is a **substantial rewording** of an idea from another text. The issue is not really how much has to be changed, but rather that the paraphrase should be **composed autonomously**.

A quotation gives the exact words of the original. It is marked by quotation marks and a reference to the original source with a page number. Use ellipsis (3 dots): … to show that something has been omitted and brackets: [added material] to show that something has been added.

Depending on the type of text cited (i.e., book, conference paper, chapter in an edited volume, journal article, website etc.) the elements of a **reference** include:
author; date; title of the book or the article; title of the journal or other work; place of publication; name of the conference; date of publication; page numbers; website address and date of access.

We cite

- to give credit to the person(s) whose work we are referring to.
- to strengthen our arguments and to make it more likely that our points will be accepted.
- to discuss current issues in the field.
- to position our work within the field.
- so that the reader can locate and read the original source.

8: AVOIDING PLAGIARISM (continued)

Unsuccessful Source Use occurs when **the relationship between the source and the new text is not transparent.**

The Consequences of Inappropriate Source Use

- The benefits of citation for the cited writer, the dissertation writer and the reader disappear.
- The result can look like plagiarism.

What You Can Do

1. Think about the ways in which your research owes a debt to other, earlier research and acknowledge those debts.
2. Select the right signals to achieve transparency of source use.
3. Be meticulous about taking notes.
4. Know *why* you want to cite each source.

9: EDITING AND REVISING

Use the following points and questions to help you revise and edit your writing.

Communication

1. Think about what your reader already knows and what you need to explain. Are your explanations clear and understandable for someone who is not familiar with your work?
2. Are there any parts of the dissertation where you should have included more information or explanation?
3. Are there any parts of the dissertation which are redundant or which should be cut down?
4. The reader may need help in order to follow the flow of your ideas. Have you previewed the contents of each chapter? Have you ended each chapter with a brief review of what you have said?

Ideas and Organisation

1. Have you introduced your subject appropriately? Have you developed it in a well-organised and logical way? Have you come to a conclusion?
2. Look at the connections in your text. Are the relationships between ideas clear and logical? Have you explicitly signalled the connections to the reader through the linking words you have chosen?

Language

1. Are there any parts of the text where you think grammar problems could interfere with understanding?
2. Are your sentences all full sentences with a subject and a verb?
3. Check that subjects and verbs agree, both singular or both plural.
4. Check your use of vocabulary. Look for unnecessary repetition, misused words, inappropriate colloquialisms and informal language.
5. Check your punctuation. Look for incomplete sentences and missing or wrongly used commas.
6. Check your paragraphing. Have you used paragraphs to break up the text according to the sequence of your ideas? Make sure paragraph breaks are at appropriate points.
7. Check for typographical errors and use a spellchecker.

10: HINTS ON WRITING

1. Plan how you will organise your dissertation. Divide each chapter into sub-sections. This will break up the writing task into parts that are short and manageable.
2. Start with the part you find easiest, perhaps a description of what you have done. Then move on to the more difficult parts as you become more familiar with your topic and experienced in writing.
3. Get your ideas down on paper first, before you revise and edit to improve the language.
4. Don't expect to get the whole dissertation perfect at the first attempt. Ideas change as your work progresses and multiple revisions are the norm in academic writing.
5. If you have a problem or a block, don't keep attempting to solve it. Rather, leave it and return to that point at a later stage or try to approach it from a different angle.
6. Print a draft version of each chapter when it is as good as you can make it. Then leave it for a time (overnight is ideal), before reading it to check for problems or parts that are unclear. Looking at your writing in print enables you to notice points that you may overlook when reading on screen.
7. When revising, the aim is to be critical, to read your dissertation as though it was the work of someone else. Of course, it is also useful to exchange dissertations with another student and get their comments.