COMPSCI 760 S1 C 2019
Machine Learning and Data Mining
Computer Science Department

Research Projects 2019

This year students will work in groups of 4 on projects. Each project will be a small research project, with a timeline is as follows:

Find a Project Group 0% Due: Friday March 8th

Topic area agreement 0% Due: Friday March 25th

Literature Survey 12% Due: Friday April 5th

Proposal area agreement 0% Due: Friday April 12th

Proposal 4% Due: Friday May 8th

Presentation 2% Due: Monday May 27th (last 5 lectures)

Final Report 20% Due: Friday June 7st

More details will be provided in individual documentations.

Peer Performance and Evaluation

In a real research projects, performance and production are required to reach the project goals (i.e. finish the paper). This is especially important in group environments where the fate of the whole is dependent on each individual.

For the proposal, literature survey and final report, each individual will submit a statement detailing each member's contributions, and this will affect the weighting of the individual marks for the team. For example if individual A was said to contribute 30% by individual B, 35% by individual C, 35% as self-assessment. Then an average (30% + 35% +35%)/3 = 33.33% weighting is given to A. This is taken into account in your overall score. If there is a high discrepancy between weighting the group will be requested to discuss this with the instructors.

The peer performance percentages will be used as a guide to the teaching staff who will make the final determination.

Sample Peer Evaluation:

Group Members (including Self)	Weighting
UPI Member 1	30%
UPI Member 2	30%
UPI Member 3	40%
Total	100%

Research Project: Topic area agreement

Worth: 0% [0 marks]

Your group must have registered as a group with Pat, Yun Sing, Joerg or Ian by the end of the day of March 29th. Each person will have the same number of groups, first come first serve.

From this papers suggested by your supervisor, you will have to branch out on your own.

Use Google Scholar to find related papers.

Research Project: Literature Survey

Worth: 12% [10 marks]

1. Summary
You need to do a literature survey on a topic concerning machine learning or data mining. The goal of the survey is to help you find a research problem that you find interesting and on which you believe you can make a contribution. To do this you should look for a topic that sounds interesting. You then need to use the resources of the internet and the library to find out what are the current problems in that area. In particular, you need to find a problem that you want to work on and start finding out what approaches have already been tried and what are their strengths and weaknesses. The weaknesses are especially important as this should give you suggestions about where you might be able to make a contribution. This will involve finding papers that describe a problem and/or approaches to solving a problem. Given a number of different approaches, you may be able to organise them along several dimensions, which can give your insight into what are possible approaches that have not yet been investigated.

2. Resources
The ACM publishes a journal called "ACM Computing Surveys" which you can access through the library's online e-journals. You go to the link to ACM Digital Library and browse the "journals" section to find the link to Computing Surveys. The important thing to remember is what you are trying to accomplish with this survey, namely, you want to end up with a problem and hopefully an idea for a new approach for your course project. The more thought you put into your exploration of the literature, the easier your life will be when doing your project!

3. What to include
Your paper should be a maximum of 10 pages long. You should include a summary of at least 9 papers (with a maximum of 30 papers). This should include papers that are recent.

Your reports may be run through Turnitin. Please refer to https://cdn.auckland.ac.nz/assets/cs/future-undergraduates/studyoptions/documents/22-Academic-Honesty-and-Special-Circumstances.pdf for details on plagiarism.

4. What to Submit

Your group should submit one literature review as a .pdf file by the due date (20 April) and each team member should also submit an individual peer performance evaluation. Please use IEEE standard paper format for your report (https://www.ieee.org/conferences_events/conferences/publishing/templates.html).

5. Marking

This is a rough guide to how the surveys will be marked. I will try to stick to the guide as much as possible but sometimes I may need to modify it for a specific case.

A. *Abstract* – (1 mark)
B. *Introduction:* Does it explain what area the literature survey is focussing on? - *(1 marks)*
C. *Problem Description:* Is the topic/problem well-defined? Is it obvious how it relates to this course? - *(2 marks)*
D. *Description of Literature:* Are the descriptions of the papers clear enough to give one an idea of what the papers' contributions are? Are there obvious gaps in the survey, e.g., are there seminal papers which should have been included? Are the papers relevant to the area? Are the papers organised along appropriate dimensions? - *(2 marks)*
E. *Summary:* Does the summary summarize where that area currently is, e.g., which problems are currently solved, which are close to being solved, which are still largely unexplored? - *(1 mark)*
F. *Citations:* Are citations used correctly, e.g., is it clear which ideas come from which paper or come from the student? - *(1 mark)*
G. *References:* Are the different papers identified sufficiently to be able to get copies? - *(1 mark)*
H. First-Cut Hypothesis: What hypothesis do you propose to test in your project? – (1 mark)

Research Project: Research Proposal

Worth: 4% [4 marks]

1. Summary

Your group needs to describe what your project is going to entail. If you are using a tool/algorithm that someone else wrote please include a citation to that, so we can find the previous research you are referring to. The same is required for datasets you will be using that come from previous work. If you are designing your own algorithm there will be less pointers to previous work, but it might be an extension of a previous algorithm, which you should cite. The most important thing is to make it as clear as possible what the project will actually entail. Please use IEEE standard paper format for your report (https://www.ieee.org/conferences_events/conferences/publishing/templates.html).

The point of the project proposal is to find out "now" if there is some "hole" in your research idea. The theory is that it is better to find this out NOW, then after you hand in your final project. To achieve this, you have to explain your project in enough detail and "clearly enough" that we can see if there are any problems with your project.

2. What to include
Your paper should be a maximum of 6 pages long.

3. What to Submit
Your group should submit one project proposal as a .pdf file by the due date (8th May) each team member should also submit an individual peer performance evaluation.

4. Marking
A. Clear definition of the problem being addressed – (1 mark)
B. Clear description of the approach taken to solve the problem – (1 mark)
C. Clear description of how the hypothesis will be tested – (1 mark)
D. Clear description of potential pitfalls which the project may encounter – (1 mark)

Research Project: Presentation

1. Summary
The main purpose of the presentations is to give each group feedback from all the supervisors. This will help them fill in holes before the research reports are due. The presentations will occur in the last two weeks of class.

2. Marking
Participating in the presentations – (1 mark)
Presenting the material in an excellent manner – (1 mark)

Research Project: Research Report

Worth: 20% [15 marks]

1. Summary
All research has a hypothesis you are trying to test. Clearly state what is the question you are trying to answer? This can be an experimental question, "will algorithm X perform better than algorithm Y on this dataset?" Or it can be about a new algorithm 'can we develop an algorithm based on "otters", Otter Colony Optimization?' Or it can even be theoretical, "can we prove when Ant Colony Optimization will find a global optima?"

Give a short description of the related research (1 page maximum!!!).

Explain how you are planning to test the hypothesis. This section will get the majority of the points! We have to understand how you went about testing your hypothesis. If it was experimental, did you use a good experimental methodology and appropriate statistics to support your results?

Describe your final results. For most of you this will include results of experiments and statistics supporting them.

Describe the future work that you think are the next steps in this work. If you were going to continue this work as a master's project, what further avenues should be explored?

Please use IEEE standard paper format for your report (https://www.ieee.org/conferences_events/conferences/publishing/templates.html). Those with good enough results might have a chance to get this work published.

2. What to include

Your paper should be a maximum of 12 pages long.

3. What to Submit

Your group should submit one final report as a .pdf file by the due date (1st June) and each team member should also submit an individual peer performance evaluation .

4. Marking

The 15 marks will be given as follows:
Description of the hypothesis you are trying to test – (3 marks)
Brief description of previous work – (2 marks)
Description of methodology used to test hypothesis or description of the case study used to demonstrate the validity of the hypothesis. The use of case study might NOT be appropriate for all project, thus approval must be obtained from the course coordinator if your team wants to conduct a case study. – (5 marks)
Description of final results- (3 marks)
Description of future work – (1 marks)
Style (if the paper looks like a research paper, with proper abstract, introduction, citations, references, tables/figures to show results, clarity in writing) – (1 mark)
