Coursework commentary 2017–18

CO3310 Artificial intelligence

Coursework assignment 2

Introduction

- Students were advised to list all references at the end of their work, which should have been properly cited whenever referred to. References to websites should include title and author if available, and date last visited.
- Answers should be expressed in the student's own words: an answer consisting entirely or mostly of quotes is unlikely to get good marks, whether or not the sources are properly referenced. Some students overlooked this, giving the bulk or at least the key parts of their answers as quotes.
- Students were also advised to submit their work as a single PDF file (**not** a zip file). Any additional files would be disregarded.

Comments on specific questions

Question 1

Machine learning and AI ethics

Part (a) was largely bookwork, and assessed students' ability to study, assimilate and explain some fairly complicated and unfamiliar material. Marks were awarded based on correctness and appropriateness of the answers as well as clarity of exposition. Good answers would have shown confident understanding of terms like "policy" and "value function" in the context of reinforcement learning, and the ways these methods learn (or do not) from outcomes of episodes and evaluations of individual states. There were some excellent answers, but also considerable variety in the depth of understanding shown.

Part (b) concerned bias in machine learning, which continues to be a contentious topic: see, for example, a recent report that Amazon had scrapped an AI recruiting tool that essentially reproduced past hiring decisions, leading to women applicants being disadvantaged. This is the kind of issue which anyone working with AI in a professional capacity may well encounter – as AI increasingly comes to permeate our daily lives, business practices and interactions with state agencies, it is essential for practitioners to keep up an informed awareness of these kinds of ethical concerns.

Marks were awarded for:

- Showing understanding of the technical, social and ethical issues discussed in the cited papers;
- Choosing appropriate examples to illustrate issues, rather than simply dealing with vague generalities;
- Structure and quality of argumentation;
- Fluency of expression.

Bolukbasi *et al* investigated some ways in which application of machine learning runs the risk of amplifying biases present in data, looking in particular at word embedding. They propose a methodology for modifying and embedding to remove gender stereotypes, such as the association between the words *receptionist* and *female*, while maintaining desired associations such as between the words *queen* and *female*, and they define metrics to quantify both direct and indirect gender biases in embeddings.

Caliskan *et al* compare word embedding techniques to Implicit Association Tests, finding stereotyped associations of "European" vs "African" American names, and conclude that learning software (LS) may perpetuate cultural stereotypes.

Wolf *et al* argue forcefully that there is always a risk that LS will behave in ways that are unexpected and undesired, since these systems are self-modifying and thus to an extent opaque to developers. This poses special imperatives for software engineering and AI ethics, especially if systems are designed to interact with humans. The development team should include experts in experimentation with human subjects. The authors argue that it is unethical to promote artificial agents or "bots" as human-like, since this raises inappropriate expectations.

There was considerable variation in the quality of answers. Some students submitted a clear, thorough and well-informed discussion of the issues and gained very good or excellent marks, while others either did not attempt the question or gave answers which were brief, unstructured, confused, poorly referenced or lacking in depth and detail.

Question 2

Planning

Part (a) was a bookwork question involving the explanation of particular terminology. Students who took the trouble to look up the definitions and paraphrase them in a way that showed good understanding could get high marks for this question, and many did get excellent or even full marks. A few did less well; for instance, circular answers like "In partial-order planning, actions are only partially ordered" are not sufficient.

Part (b) involved PDDL and required students to write a plan for a "toy" problem using their own action and state definitions.

Answers to this kind of question will need to spell out any assumptions that are made, for example, whether the travellers are already at the respective airports. It can be assumed that flights are instantaneous, and that the plan generator has access to airline timetables. There are various ways to address this question, but the hard work is in setting up the state and action definitions. Too many answers used highly specialised predicates/actions like TravelTogether, ArriveBefore … and/or included the particulars of the problem in what should be generic action definitions. There should be some explanation (which was rare) of how a constant term identifies a unique flight.

Some students gave very good or excellent answers that fulfilled all the requirements. Several students either did not attempt this question, or lost marks for one or more of the following reasons:

- Not giving action/state definition as required.
- Inadequate explanation.
- Incomplete solutions to the problem.
- Over-specialised/ad hoc definitions as noted above.
- Using terms (*e.g.* 'SameFlight') that were not explicitly defined.

- Not adhering to PDDL syntax.
- Omitting key steps in the plan, e.g. boarding a plane.

Question 3

Natural language processing

For part (a), students were required to show one parse tree for each sentence if possible (even if there were multiple parses) and if not, to explain why a sequence could not be parsed according to the grammar. This question is fairly straightforward and many students got excellent or full marks; some lost marks by failing to explain their answers, while a few showed poor understanding of analysing and representing syntactic structure.

Part (b) required students to extend the coverage of the grammar from (a). Note that it was neither necessary nor sufficient to give a full set of rules; students should explain which constructions need to be covered, and state which new proposed rules will take care of these constructions. Marks were awarded for correctness of the proposed rules and appropriate explanations, as well as conciseness and generality. For example, if the data includes a sequence of two adjectives it is preferable to code rules that generate an indefinitely long series of adjectives rather than precisely two.

Some students submitted "ad hoc" grammars which are tailored to the particular strings in the examples but do not respect linguistic structure, where, for example, S nodes dominate sequences which are not sentences or clauses.

A few students got very good or excellent marks for this question; several others did not attempt it, submitted an incomplete answer, or lost marks by giving inadequate explanations and/or coding "ad hoc" grammars which would not generalise well to novel sentences.

Part (c) required students to extend the augmented grammar on p.45 of the subject guide to include verbs with three arguments such as *give*(a, b, c), meaning "a gives b to c", and to construct augmented parse trees for both examples below.

1. John gives Mary a wumpus.
2. John gives a wumpus to Mary.

One problem here is that the arguments come in different orders in the two sentences, so the grammar will need separate lexical rules for both variants. Students tackling this exercise may well have concluded that the original grammar was over-simplified and can't easily be scaled up without either adding more complex rules or cutting corners. Credit was given for any reasonable attempt. It is advisable to show a full derivation to be sure that the rules actually work; in particular it should be made clear how the two different orders of arguments map onto the same logical statement. In many cases it was not at all clear how the derivation was supposed to work.

This is a relatively challenging exercise and was marked charitably, though for full marks students must give a derivation which actually matches their grammar. Good answers at least extended the formalism to handle 3 arguments rather than 2. Very good/excellent answers included separate versions of rules to handle the different orders of arguments (without simply assuming that expressions can be applied in an arbitrary order). Outstanding answers would have coded the grammar to select the appropriate semantics for each construct. Several students did not attempt this question, but most of those who did got at least 'Good' marks with some achieving 'Very good/ excellent'. A few students scored poorly by submitting incomplete or unclear answers which showed inadequate understanding of the problem.