

Assignment 2

Marking [100%] due Fri 08 Apr 2022 at 16:00 (UK time)

60%: Project report

10%: Colab notebook to demo your project

10%: Short 5min recording of a spotlight talk about your project

20%: Individual peer feedback

Project Report

For this assignment, work as a group on a machine learning project in Natural Language Processing. You will prepare a report (maximum eight pages; excluding references) that delivers (i) an introduction including a well defined research hypothesis, (ii) a literature review, critically discussing work related to your project, (iii) a methods section describing your model/approach, (iv) a clear description of your experiments, (v) a critical discussion of your findings including a quantitative and qualitative analysis of your empirical results, and (vi) a conclusion and description of future work.

You can choose to either work on one of the default projects that we outline below or to propose your own project. We highly encourage you to do the latter and to propose your own project. Your project should (a) have a strong machine learning or deep learning component and (b) make extensive use of natural language data.

Submit a report as a single PDF with a maximum of eight pages (excluding references and an optional appendix). The report should at least include the following (for more details see the marking scheme appended to this document):


- An abstract briefly summarizing the research problem and the outcome of your work
- An introduction describing the problem/research question on a high-level as well as the main outcomes of your work.
- A related work section discussing how what you did relates to prior work.
- A methods section, detailing what you did, how you did it and with what motivation. Make sure to formalize your models.
- An experiments section, describing the empirical setup of your research and metrics.
- A results and discussion section, presenting your findings as well as critically analyzing your findings.
- A conclusions section, summarizing your research outcome, as well as discussion of future work if you or others were to build upon your work.

For advice on writing a good report, see [1] and [2].

Talk Recording and Demo Notebook

In addition to the project report, we ask you to prepare and submit a recording of a short 5min spotlight talk about your project. For the talk, prepare a few slides that outline the problem setting, your approach and results. Furthermore, we ask to demonstrate that your findings are reproducible. To this end, submit a single zip file containing the source code of your project. In addition to the zip file, prepare a self-contained notebook demonstrating your approach so that it can be uploaded and executed on Colab. If your project is less about a predictive model but more about an explorative or descriptive research, the notebook should demonstrate your finding. If your project has a stronger machine learning component then please avoid training a model in the notebook. Instead, give us an **easy way** to test your system on unseen inputs. In case you need to train models, do this training offline and load the model into Colab.

Groups

Please form groups of exactly four students and add your group to  Assignment Groups 2022. If you have difficulties finding a fourth member, add your group to the corresponding column “We are looking for a fourth member” in the “Blackboard” tab. Similarly, if you are looking for a group, add yourself to the corresponding column “I am looking for a group” in the “Blackboard” tab.

We are only accepting submissions from groups with exactly four students. If there are multiple three student groups by the end of February, we will randomly split and reassign these groups.

Ethics and Project Feedback

Before you commit to a project, please send a 100 word abstract of your proposed work in private to the lecturers and TAs on Moodle via the “Discuss with TA/LecturerForum” area. This will help us to understand and approve the ethical implications of your projects, and give light feedback if we see any obvious practical pitfalls. To this end, please include a description of the data you plan to use. We aim to give feedback within three days (excluding weekends). Note that grading of the final project will depend on execution of the project idea only, not on the idea itself or our approval thereof.

Default Projects

Develop a new model for one of the challenges below:

- Cross-lingual Extractive Question Answering
 - Dataset: <https://github.com/facebookresearch/MLQA>
 - Paper: <https://arxiv.org/abs/1910.07475>
- CoNaLa challenge
 - Dataset: <https://conala-corpus.github.io/>
 - Paper: <https://arxiv.org/abs/1805.08949>
- WikiHow Text Summarization
 - Dataset: <https://github.com/mahnazkoupaei/WikiHow-Dataset>
 - Paper <https://arxiv.org/abs/1810.09305>

- Utilize the ATlas Of MachIne Commonsense (ATOMIC) dataset of commonsense knowledge for a downstream NLP task
 - Dataset: <https://homes.cs.washington.edu/~msap/atomic/>
 - Paper: <https://arxiv.org/abs/1811.00146>
- WinoGrande
 - Dataset: <http://winogrande.allenai.org/>
 - Paper: <https://arxiv.org/abs/1907.10641>

Options for Computational Resources

- Colab (see <https://research.google.com/colaboratory/faq.html#gpu-availability>)
 - You get one free GPU per person for 12 consecutive hours
 - 4 people = 4 GPUs
 - You could train for longer than 12 hours if you are clever about checkpointing and regular saving your models
- UCL Cluster

Style Guides and Submission

You are required to use the ACL 2021 style template for your report. You can either use LaTeX or Word (<https://2021.aclweb.org/downloads/acl-ijcnlp2021-templates.zip>). Submit a single PDF of eight pages maximum (excluding references and an optional appendix) to Moodle. Use only your group ID as the name (e.g. 123.pdf). Do not change the template (e.g. by reducing or increasing the font size, margins etc.).

Marking and Peer Feedback

Your report will be assessed based on the criteria above. In addition to your group project, we ask you to anonymously comment on your collaborators regarding the quality of their contribution to your joint project. We will release a survey to submit your feedback closer to the assignment deadline, but the questions are going to roughly match the template from JHU that is attached to this assignment. Peer feedback will make up 20% of a student's mark in terms of their individual performance in the group project. The marking criteria for the report of this assignment follow those of the MSc project criteria and can be found below.

Marking Descriptors

90-100% Exceptional (Distinction)

- A clear contribution to the field, of publishable quality, excellent report
- Evidence of considerable extra-curricular reading, critical thought and original interpretation
- Challenging goals have been fully met, substantial deliverables, research level insight needed
- Close to faultless in execution and write-up, a high level of independence

This represents a really outstanding achievement. The project needs to clearly stand out above others. A mark in this range is hard to achieve and rare (< 1%).

80-89% Outstanding (Distinction)

- Potential contribution to the field, could lead on to publishable work, very good report
- Evidence of significant extra-curricular academic, critical thought and original interpretation
- Goals met, only very minor faults in execution, depth of understanding or write-up
- Challenging project and substantial deliverables, largely self-directed

Excellent in most respects but doesn't fully meet the criteria for the top range. A small number of projects are in this range each year (2-3%).

70-79% Excellent (Distinction)

- Very well written report with a clear logical structure
- Good demonstration of critical thought, understanding and extra-curricular reading
- Some minor faults in execution or understanding, otherwise carried out effectively, most or all goals fully achieved
- A good level of challenge, substantial deliverables, and a good level of self-direction

This represents a straightforward distinction project. Most things have been done well, but there will be some faults or criticisms. The goals have been met. A reasonable number of projects can be expected to achieve this level (~20%).

60-69% Good (Merit)

- Clear project write-up with logical structure
- Evidence of understanding, some extra-curricular reading, and critical thought
- May contain some ambiguities or faults, not all goals fully achieved
- Reasonable level of challenge, good quality deliverables, satisfactory self-management, with some supervision help needed

A good result, that is well on the way to delivering most features, but is not fully complete or finished, or has a lower level of challenge. The majority of projects are likely to be at this level.

50-59% Satisfactory (Pass)

- Adequate project write-up, lacking clarity or detail in places, or containing irrelevant material
- Limited evidence of extra-curricular reading or original thought, mostly demonstrates understanding of core issues
- Some significant deficiencies or incomplete goals, deliverables adequate but of limited quality
- Project not particularly ambitious or challenging, more significant supervision help required

A satisfactory but limited result. The core features are in place but may be buggy or not that well defined. Enough has been done to present a viable solution, of which at least some parts can be demonstrated. A minority of projects in this range.

45-49% Borderline fail (Could pass with extra work)

- Write-up is sub-standard, with noticeable errors or omissions, but could be made passable within a reasonable time
- Some clear flaws in understanding, limited or no extra-curricular reading
- Actual achievements not very substantial or challenging, deliverables of lower quality or incomplete, but could be improved fairly easily
- Not quite enough challenge or depth demonstrated, required significant extra supervision or there was a failure to attend tutorials

The project has enough substance to demonstrate it could be made into a pass in a relatively short length of time but is still missing significant features, or the write-up fails to describe what was actually achieved.

0-44% Unsatisfactory (Clear fail)

- Write-up is insubstantial, incoherent, rushed, has important omissions, or irrelevant material
- Serious flaws in understanding, little or no extra-curricular reading
- A lack of concrete achievements, substantial parts missing, few deliverables
- Serious lack of challenge or depth demonstrated, lack of engagement, required excessive supervision or there was a failure to attend tutorials

The basis of a viable project may be present but is a long way from being completed. A significant amount of additional work would be needed to reach a passable standard.

References

- [1] <http://approximatelycorrect.com/2018/01/29/heuristics-technical-scientific-writing-machine-learning-perspective/>
- [2] <https://naacl2018.wordpress.com/2017/12/01/musings-on-writing-on-an-nlp-cl-paper/>
- [3] <https://guides.nyu.edu/c.php?g=276826&p=1846154>

[4] <https://ethics.grad.ucl.ac.uk/>

[5]

https://www.dhs.gov/sites/default/files/publications/CSD-MenloPrinciplesCORE-20120803_1.pdf

[6] <https://www.ucl.ac.uk/research/integrity/sensitive-research>

Peer Evaluation Form for Group Work

Adapted from a peer evaluation form developed at Johns Hopkins University (October, 2006)

Your name _____

Write the name of each of your group members in a separate column. For each person, indicate the extent to which you agree with the statement on the left, using a scale of 1-4 (1=strongly disagree; 2=disagree; 3=agree; 4=strongly agree). Total the numbers in each column.

Evaluation Criteria	Group member:	Group member:	Group member:	
Attends group meetings regularly and arrives on time.				
Contributes meaningfully to group discussions.				
Completes group assignments on time.				
Prepares work in a quality manner.				
Demonstrates a cooperative and supportive attitude.				
Contributes significantly to the success of the project.				
TOTALS				

