Project 2: Neural Machine Translation

April 26, 2018

This project will help you
In summary, your task is to

- Implement

1 Neural Machine Translation

1. a) Implement

2 Data

All relevant data (including details about file formats) are available from https://uva-slpl.github.io/nlp2/projects.html.

In this project, you will work with a parallel corpus taken from

We are making available *training* data (which you can use to perform parameter estimation), *validation* data (which you can use to debug your implementation as well as to perform model selection), and finally in due time *test* data (which you will use to conduct your final empirical comparison).

3 Report

You should use latex for your report, and you should use the ACL template available from http://acl2017.org/downloads/acl17-latex.zip (unlike the template suggests, your submission should not be anonymous).

We expect short reports (4 pages plus references) written in English. The typical submission is organised as follows:

- abstract: conveys scope and contributions;
- introduction: present the problem and relevant background;

- model: technical description of models;
- experiments: details about the data, experimental setup and findings;
- conclusion: a critical take on contributions and limitations.

4 Submission

You should submit a tgz file containing a folder (folder name `lastname1.lastname2`) with the following content:

- Test predictions (in naacl format) using your best run for each of the following models
  - –
- Report as a single pdf file (filename: `report.pdf`)

Your report may contain a link to an open-source repository (such as github), but please do not attach code or additional data to your tgz submission.

You can complete your project submission on Blackboard.

5 Assessment

Your report will be assessed by two independent reviewers according to the following evaluation criteria:

1. Scope (max 2 points): Is the problem well presented? Do students understand the challenges/contributions?
2. Theoretical description (max 3 points): Are the models presented clearly and correctly?
3. Empirical evaluation (max 3 points): Is the experimental setup sound/convincing? Are experimental findings presented in an organised and effective manner?
4. Writing style (max 2 points): use of latex, structure of report, use of tables/figures/plots, command of English.
5. Extra.