COMP7607: Natural Language Processing

Project Presentation | A2 FAQ | A3 Release

Yiheng Xu

Email: yihengxu@connect.hku.hk

Project Proposal (5%)

- A 1-2 page proposal
- Most proposals are really professional:
  - Paper Choice&Study: a good start point with novel ideas
  - Project Plan: timeline table, gantt chart, etc
  - Format: LaTeX with academic template, good references

Project Proposal (5%)

- A 1-2 page proposal
- Some minor problems:
  - Paper choice:
    - Propose a project without survey and estimation (pretraining) X
    - Properly estimate required resources (fine-tuning provided model is okay!) ✓
  - Contribution:
    - Simply indicate equal contribution X
    - Indicate the detailed contribution of each individual in different aspects✓

Project Presentation (10%)

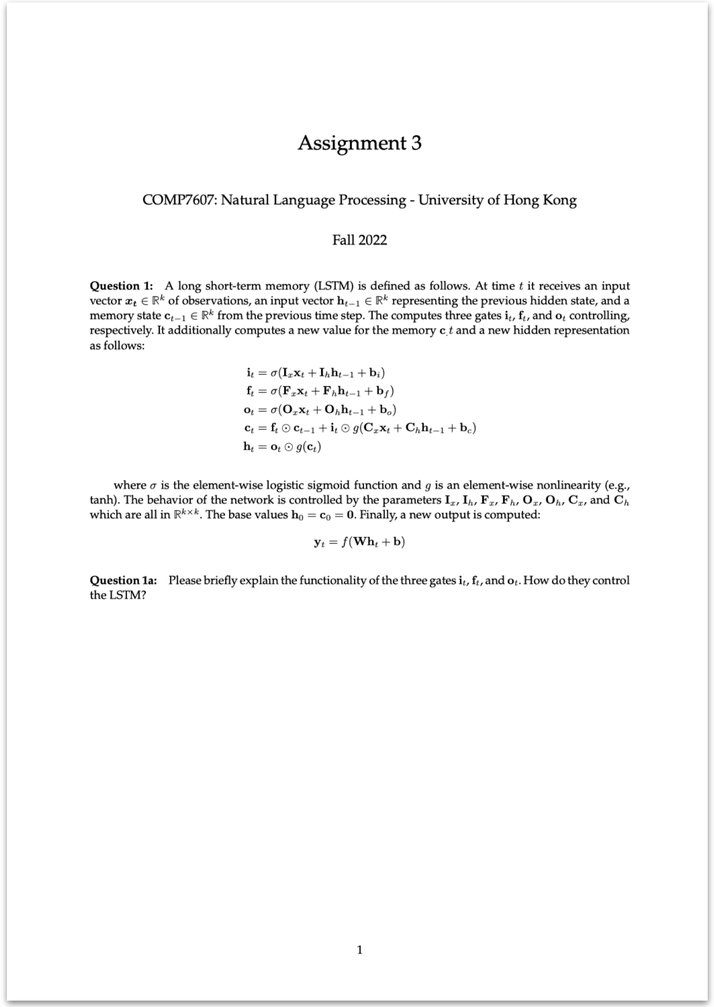
- A **5-minute (strict limit)** midway presentation with a maximum of 5 slides (no strict limit)
- 32 groups in 3 Sessions (All team members should attend your session in person.)
  - 19:00-20:00 Group 1-10
  - 20:00-21:00 Group 11-21
  - 21:00-22:00 Group 22-32
- You presentation will be peer-reviewed
  - Professor & TA.
  - Groups in same session.

Assignment 2 FAQ

- Q: Implement both taggers?
  - Yes.
- Q: Tensorflow/Keras?
  - Yes
- Q: About the output file test.txt.
  - Handbook: “your submission should be in the same format as the test.txt file, but replace the ground truth labels with your prediction.”
  - Keep the prediction in the same order: torch.utils.data.SequentialSampler

Assignment 3 Release

- 7 questions. No coding.
- Deadline: Nov 30 11:59 pm



Assignment 3

COMP7607: Natural Language Processing - University of Hong Kong

Fall 2022

Question 1: A long short-term memory (LSTM) is defined as follows. At time t it receives an input vector $\boldsymbol{x_t} \in \mathbb{R}^k$ of observations, an input vector $\mathbf{h}_{t-1} \in \mathbb{R}^k$ representing the previous hidden state, and a memory state $\mathbf{c}_{t-1} \in \mathbb{R}^k$ from the previous time step. The computes three gates $\mathbf{i}_t$, $\mathbf{f}_t$, and $\mathbf{o}_t$ controlling, respectively. It additionally computes a new value for the memory $\mathbf{c}\,t$ and a new hidden representation as follows:

$$\mathbf{i}_t = \sigma(\mathbf{I}_x\mathbf{x}_t + \mathbf{I}_h\mathbf{h}_{t-1} + \mathbf{b}_i)$$
$$\mathbf{f}_t = \sigma(\mathbf{F}_x\mathbf{x}_t + \mathbf{F}_h\mathbf{h}_{t-1} + \mathbf{b}_f)$$
$$\mathbf{o}_t = \sigma(\mathbf{O}_x\mathbf{x}_t + \mathbf{O}_h\mathbf{h}_{t-1} + \mathbf{b}_o)$$
$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot g(\mathbf{C}_x\mathbf{x}_t + \mathbf{C}_h\mathbf{h}_{t-1} + \mathbf{b}_c)$$
$$\mathbf{h}_t = \mathbf{o}_t \odot g(\mathbf{c}_t)$$

where σ is the element-wise logistic sigmoid function and g is an element-wise nonlinearity (e.g., tanh). The behavior of the network is controlled by the parameters $\mathbf{I}_x$, $\mathbf{I}_h$, $\mathbf{F}_x$, $\mathbf{F}_h$, $\mathbf{O}_x$, $\mathbf{O}_h$, $\mathbf{C}_x$, and $\mathbf{C}_h$ which are all in $\mathbb{R}^{k\times k}$. The base values $\mathbf{h}_0 = \mathbf{c}_0 = \mathbf{0}$. Finally, a new output is computed:

$$\mathbf{y}_t = f(\mathbf{W}\mathbf{h}_t + \mathbf{b})$$

Question 1a: Please briefly explain the functionality of the three gates $\mathbf{i}_t$, $\mathbf{f}_t$, and $\mathbf{o}_t$. How do they control the LSTM?

1