Deep Learning for Natural Language Processing

Exercise 1 – Kick-off

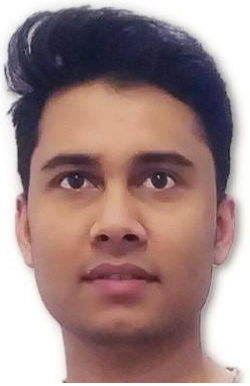
Niraj Dev Pandey

Natural Language Learning Group
Technische Universität Darmstadt

Who am I? – Niraj Dev Pandey

TECHNISCHE UNIVERSITÄT DARMSTADT

- M.Sc. Data Analytics & Machine Learning
  - University of Hildesheim, Germany
  - Information Systems and Machine Learning Lab (ISMLL)



Today

TECHNISCHE UNIVERSITÄT DARMSTADT

- Pingo: warm-up questions
- Orga
- Python, numpy and Docker

Warm-up questions

TECHNISCHE UNIVERSITÄT DARMSTADT

- Please navigate to...

https://

pingo.coactum.de/806359

Today

TECHNISCHE UNIVERSITÄT DARMSTADT

- ✓ Pingo: warm-up questions
- Orga
- Python, numpy and Docker

Practice Classes

TECHNISCHE UNIVERSITÄT DARMSTADT

- **When?**
  - Tuesdays
  - If inverted classroom: Online
  - Else: **15:20-17:00** in **S207/167**
- **What?**
  - We will discuss and implement topics and techniques introduced in the lecture
- **Where from?**
  - Moodle: https://moodle.informatik.tu-darmstadt.de/course/view.php?id=587

Practice Classes: General Structure

TECHNISCHE UNIVERSITÄT DARMSTADT

1. Warm-up questions (Pingo)
2. Exercises
   - Alone or in groups
   - Using a calculator and/or a laptop
3. Discussion of your solutions
4. Introduction to the new home exercise
5. If necessary and still time: discussion of last home exercise

Home Exercises

TECHNISCHE UNIVERSITÄT DARMSTADT

- Content of home exercises:
  - Mostly programming
  - Sometimes math
  - Answering question to a given paper (relevant for the exam!)
- There will be 10 home exercises, worth ~100 points in total
- Schedule:
  - Home exercise released in Moodle in week *i*
  - Hand in **until Monday, 13:00** in week *i+1*

Home Exercises (cont.)

TECHNISCHE UNIVERSITÄT DARMSTADT

- Submission format:
  - Python implementations: as plain python scripts (**.py**)
    - **Must** be runnable in the Docker container provided by us (details in a minute)
    - Please provide some comments and/or documentation
  - Your answers to questions, discussions, etc.: as a **PDF** file
  - In the case of multiple files: submit one **zip-archive**
- Will it be allowed to submit in groups?
  - No.
- Plagiarism won‘t be tolerated. See rules at https://www.informatik.tu-darmstadt.de/de/studierende/studium/plagiarismus/

Exercises – Why bother?

TECHNISCHE UNIVERSITÄT DARMSTADT

- They‘re relevant for understanding the subject matter of the lecture
- They‘re relevant for the exam.
- To enable you to apply deep learning...
  - ...for your own project ideas
  - ...in your thesis
  - ...at a company after graduating
- And...

Exam Bonus

TECHNISCHE UNIVERSITÄT DARMSTADT

- By participating in the practice class, you can get a bonus of 0.3 (or 0.4) for the exam – Or even a full grade
- The following rules apply:
  - You need to reach 70 out of 100 points for the home exercises
  - ~10 homeworks 10 points each

1. You need to participate in a “shared task”
   - Will take place in the 2nd half of the course
2. You have to pass the exam without the bonus, i.e. the bonus cannot turn a 5 into a 4
3. The bonus is valid until the next term in which the lecture + exercise takes place == this semester and next semester
4. The bonus can only be used once per student

Mock Exam

TECHNISCHE UNIVERSITÄT DARMSTADT

- A mock exam will replace the lecture and exercise on **25.06.20**

Shared Task

TECHNISCHE UNIVERSITÄT DARMSTADT

As already mentioned, there will also be a **Shared Task.**

Work on a common (real NLP) problem!

This year: **poetry generation**

- You will work in groups of up to 3 people
- You will develop a solution to the problem and then compete with other groups

Shared Task (cont.)

TECHNISCHE UNIVERSITÄT DARMSTADT

- Grading: up to 100 points can be reached. Grading is based on:
  - A written report about your system (60 points)
  - Your ranking compared to other teams -> better system, more points (30 points)
  - A short presentation in the last class (10 points)
  - **The points you earn in the shared task will count towards the points necessary to obtain the bonus**

- Planned kick-off: **14.05.2020**
- Submission: **~18.07.2020**

A Look back at the Shared Task 2017

TECHNISCHE UNIVERSITÄT DARMSTADT

- The DL4NLP2017 shared task deals with classification of short text snippets into one of three classes, namely, MATERIAL, TASK, and PROCESS

- Such classification may be useful for the analysis of scientific texts

A Look back at the Shared Task 2017 (cont.)

TECHNISCHE UNIVERSITÄT DARMSTADT

- Each text snippet, such as

  Transmission Electron Microscope

  is already extracted

- Along with the text snippet, we provide the (1) **text itself** from which it was extracted as well as the (2) **snippet's offset** in the text

A Look back at the Shared Task 2017 (cont.)

TECHNISCHE UNIVERSITÄT DARMSTADT

For example, the surrounding left context of *Transmission Electron Microscope* is:

[...] *information gathered using EDS. Cross-sections and*

Its surrounding right context is

(TEM) samples were produced using a dual beam FEI [...]

You are free to use this or other context to classify *Transmission Electron Microscope* into its gold class, which is MATERIAL in this case.

In the case of questions…

TECHNISCHE UNIVERSITÄT DARMSTADT

- **Tutor hours:**
  - Held by Niraj Dev Pandey: deeplearning4nlp@gmail.com
  - Purpose: questions on Docker, python, Tensorflow, Keras
  - Contact via Moodle messages
  - Monday, 12:00-12:30 on Skype group

- **Questions about the lecture or the exercises?**
  - please post them on the Moodle forum.

  Questions about home exercise grading?
  - please contact the tutors via Moodle

Today

TECHNISCHE UNIVERSITÄT DARMSTADT

- ✓ Pingo: warm-up questions
- ✓ Orga
- Python, numpy and Docker

Today

TECHNISCHE UNIVERSITÄT DARMSTADT

- ✓ Pingo: warm-up questions
- ✓ Orga
- ✓ Python, numpy and Docker

- **Make sure to bring a calculator / laptop for the next exercise.**