UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

COS711 Assignment 3

Deep Learning

Due date: 29 October 2019, at 16:30

1 General instructions

For this assignment, you will use deep learning techniques to solve a real life problem. You will have a choice between two problems hosted on `https://zindi.africa`, an African data science and machine learning competition platform. Based on the problem you choose, you may be allowed to work in a group of two. You will be required to demonstrate your work to the lecturer and fellow students, explaining the design choices you have made while constructing your deep learning models. The assignment is out of 100 marks in total.

2 Deep Learning

Deep learning is a term used to refer to a modern branch on neural network models focused on learning hierarchical representations of the data rather than modelling a single highly complex non-linear transformation from the inputs to the outputs. Deep learning architectures such as convolutional neural networks (CNNs) and long short-term memory networks (LSTMs) have achieved impressive results in image processing and natural language processing tasks. For this assignment, you will apply two or more deep learning techniques of your choice to a real-life problem.

2.1 Option 1: Soil Moisture Prediction

For this task, you will need to create a deep learning solution to predict soil moisture based on past data. Thus, this problem is of temporal nature. Full details of the problem, as well as the dataset, are available here:
`https://zindi.africa/competitions/wazihub-soil-moisture-prediction-challenge`

2.1.1 Problem description

> *"In the face of climate change, the agricultural sector in Africa needs to adapt its practices. Being able to accurately measure and predict soil humidity in their fields will allow farmers to prepare their irrigation schedules optimally and efficiently.*
>
> *Sensor-based irrigation and machine learning algorithms can provide farmers with a solution to manage water usage more efficiently. However, current machine learning algorithms built on sensor data require a lot of data for proper training. Stable sensor data is difficult to obtain in rural Africa where many problems arise such as accessibility, limited battery power, lack of internet, humidity/heat problem.*

The objective of this Zindi competition is to create a machine learning model capable of predicting the humidity for a particular plot in the next few days, using data from the past. A part of the challenge is to design algorithms that are resilient and can be trained with incomplete data (e.g. missing data points) and unclean data (e.g. lot of outliers).

This resulting model will enable farmers to anticipate water needs and prepare their irrigation schedules."

2.1.2 Your task

Your task is to apply deep learning techniques of your choice to create a soil moisture prediction model. You must apply at least two different deep learning architectures. You may, for example, use a generative model to generate more training data, and then use a convolutional/LSTM model to learn from the data. Alternatively, you may compare two different deep learning models on the given data set. The more different the two techniques you pick are, the higher mark you will get. You are allowed to use more than two deep learning techniques if you wish.

2.1.3 Notes

- You may (but don't have to) work in a group of 2 on the soil moisture prediction problem.

2.2 Option 2: Pothole Image Classification

For this task, you will need to create a deep learning solution to detect potholes in the road images captured by a dash camera. Full details of the problem, as well as the dataset, are available here:
`https://zindi.africa/competitions/miia-pothole-image-classification-challenge`

2.2.1 Problem description

"Potholes have become a huge problem for most drivers. With the South African government spending over R22 billion over the past 3 years on pothole repair programs and the Automobile Association (AA) acknowledging more than 5% of road deaths to unmaintained road structure (potholes).

The objective of this challenge is to create a machine learning model to accurately predict the likelihood that an image contains a pothole."

2.2.2 Your task

Your task is to apply deep learning techniques of your choice to create a pothole detection model. You must apply at least two different deep learning architectures. You are also required to optimise the proposed architectures, and compare at least two different data augmentation and/or transformation techniques.

2.2.3 Notes

- You must work on the pothole image classification individually: no group effort is allowed.

3 General Notes

- Implementation
  - You may use any programming language and platform
  - You may use a neural network library/framework
  - You must submit a zip file with your code together with a short description to the CS website.
- Demo
  - Assignments will be marked via demos in class
  - Each demo has a hard time limit of **10 minutes**.
  - You must specify what each team member contributed
  - Refer to the next section for a mark breakdown

4 Demo and Marking

The following points have to be addressed in the demo (marking allocation shown in square brackets):

1. Data preprocessing and experimental set-up **[10%]**
   (a) What transformations, if any, have you applied to the data? What part of the data set did you use for training and testing?
   (b) What software tools have you used for data processing?
2. Deep learning models **[30%]**
   (a) What deep learning techniques have you employed? Why did you choose these techniques?
   (b) What parameter optimisation have you performed? How did you choose the network architectures?
   (c) What software tools have you used for the deep learning models?
3. Training algorithm **[10%]**
   (a) What training algorithm did you use? Why?
   (b) How did you optimise the algorithm parameters?
   (c) What software tools and statistical methods have you used for training/evaluation?
4. Results **[50%]**
   (a) You will be asked to demonstrate your model.
   (b) Discuss the results: provide at least the training, generalisation, and classification errors. Errors should be reported as averages over multiple independent simulations, and the corresponding standard deviations must be provided. Any additional model quality measures can be added at your discretion.
   (c) Discuss possible reasons for the observed performance.
   (d) Discuss the challenges you have faced during the completion of the assignment. What proved to be the largest road block?