Assignment 5, Project proposal

Philip Bouman (10668667)

Overview

The goal of the project is to find relations between characters of different types of writing systems (scripts). In the first part of the research a deep convolutional neural network will be trained to classify (images of) characters. A deep convolutional neural network is a machine learning architecture consisting of (multiple) stacks of layers. Depending on the availability of (training)data the training/learning process can either be supervised, using phonemes to label characters or unsupervised, using just the characters. The second part of the project will focus on getting a better understanding of the representations learned by the network. This will mainly be done by investigating and visualizing features learned in individual layers and smaller combinations of layers. The results can be compared to existing image recognition networks or between methods (for example if supervised and unsupervised are both used).

Literature review

Convolutional neural networks (CNN) are among the most suitable architectures for handwritten digit and character recognition[3]. Neural networks learn intermediate representations of their inputs, which can be useful for subsequent classification tasks. The structure of CNNs, mimicking receptive fields, followed by pooling and activation operations is particularly useful for handling data that can be represented as two- or higher-dimensional input, such as images[4]. Deep CNNs also incorporate many layers and thus many intermediate representations, while keeping the number of free parameters small[2]. Recent work with CNNs on computer vision was applied to recognition of natural images and 3D objects, image denoising and image segmentation. CNNs seem to work well for supervised and unsupervised learning tasks and can be succesful using relatively small datasets[3].
Orthography is the set of conventions for writing a language, such as spelling, hyphenation and punctuation. Orthography is also largely concerned with the relation between phonemes and graphemes. Graphemes represent the smallest individual characters of a writing system and phonemes relate to the sound of particular graphemes[5]. Phonemes represent a standardized collection of spoken language and are collected in the International Phonetic Alphabet (IPA)[1].

Graphemes denote in principle the phonemes of a language, since writing systems emerged from spoken language. This research will focus on exploring the grapheme-phoneme relationship the other way around, taking graphemes from different writing systems and classifying them to phonemes. Is the manner in which characters are written any indication of the sound they will represent?

Research question

Can a CNN succefully be used to classify graphemes of different writing systems to related general phonemes?

Method and approach

The first step of the project is collecting more data and converting it to the proper file format. The current dataset needs to be extended with more writing systems, as to have enough data to train the CNN. Right now it looks like this can be done using a webcrawl of Scriptsource[1]. The collected scripts will be digital fonts using the .ttf or .otf file format. To easily work with the data it has to be converted to images (PNG or JPEG). This can be done using the Python Image Library (PILLOW). After the conversion to image format, either features need to be extracted using seperate software (for instance OpenCV), or the image can be feed directly to a machine learning library, which can extract feautures itself. For creating the architecture of the CNN an existing machinle learning API will be used. This is most likely gonna be Keras, which is a high-level overlay API for either the TensorFlow or Theano library. Keras allows for fast prototyping, modularity and extensibility especially for CNNs. Next the architecture needs to be trained on part of the data and tested on another part of the data. In turn this will lead to a recurrent process of altering the architecture, tweaking the parameters and training again to achieve desired results.

Evaluation

The evaluation will be done using F-measure scores on the test data. These scores can be compared to results of similar computer vision networks. Another part of the evaluation will consist of investigating the feature representations learned by the network. By visualising these representations they can be compared to other research concerning shape-sound relations.

[1] http://scriptsource.org/cms/scripts/page.php

Plan

See chart. I divided the planning in 3 global components (each with smaller subtasks):

- preparation (4 weeks)
- implementation (5 weeks)
- evaluation & report (3 weeks)

References

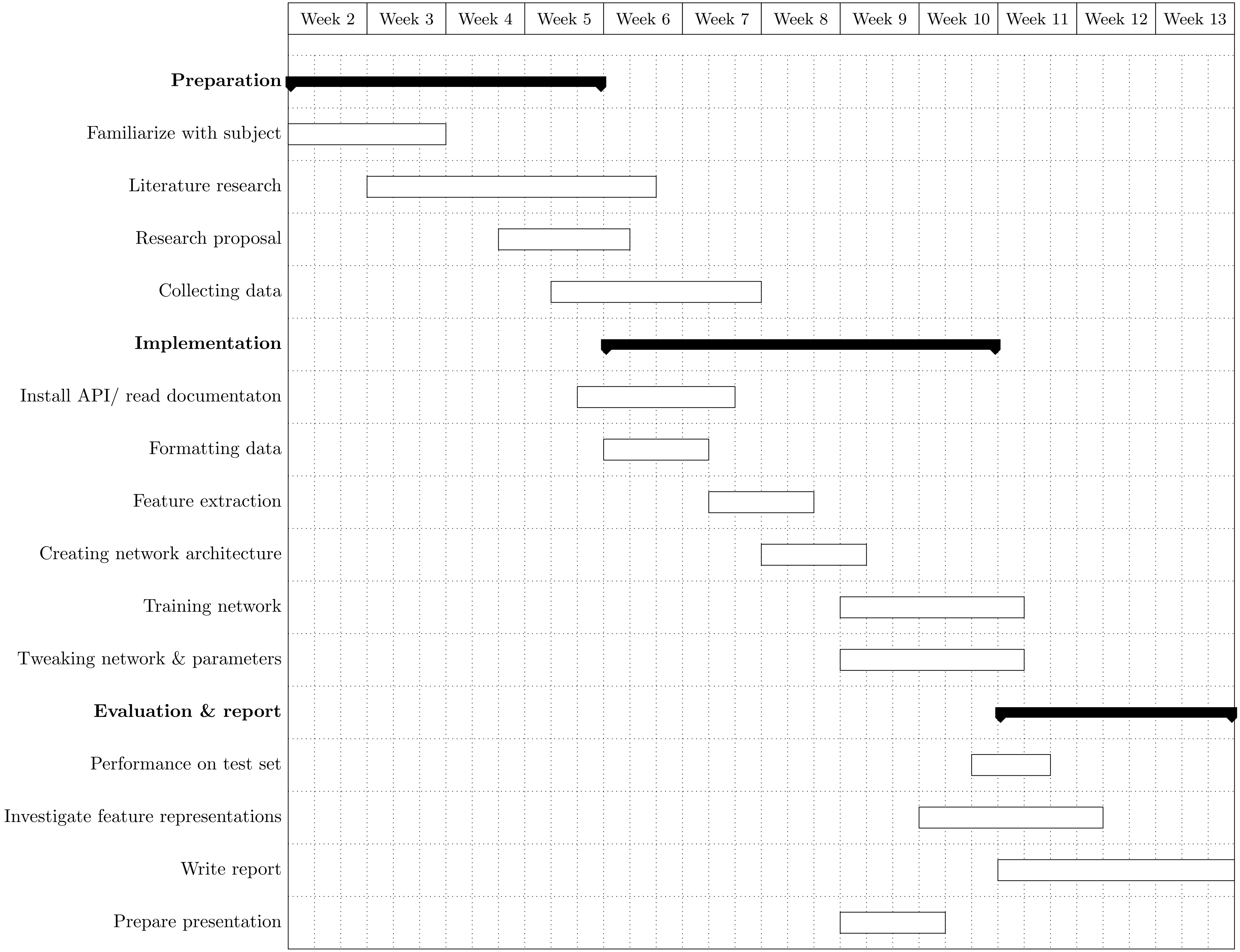
[1] International Phonetic Association. *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press, 1999.

[2] Théodore Bluche, Hermann Ney, and Christopher Kermorvant. Feature extraction with convolutional neural networks for handwritten word recognition. In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 285–289. IEEE, 2013.

[3] Dan Claudiu Ciresan, Ueli Meier, Luca Maria Gambardella, and Jurgen Schmidhuber. Convolutional neural network committees for handwritten character classification. In *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, pages 1135–1139. IEEE, 2011.

[4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[5] Mark S Seidenberg. Beyond orthographic depth in reading: Equitable division of labor. *Advances in psychology*, 94:85–118, 1992.


Week 2
Week 3
Week 4
Week 5
Week 6
Week 7
Week 8
Week 9
Week 10
Week 11
Week 12
Week 13
Preparation
Familiarize with subject
Literature research
Research proposal
Collecting data
Implementation
Install API/ read documentaton
Formatting data
Feature extraction
Creating network architecture
Training network
Tweaking network & parameters
Evaluation & report
Performance on test set
Investigate feature representations
Write report
Prepare presentation