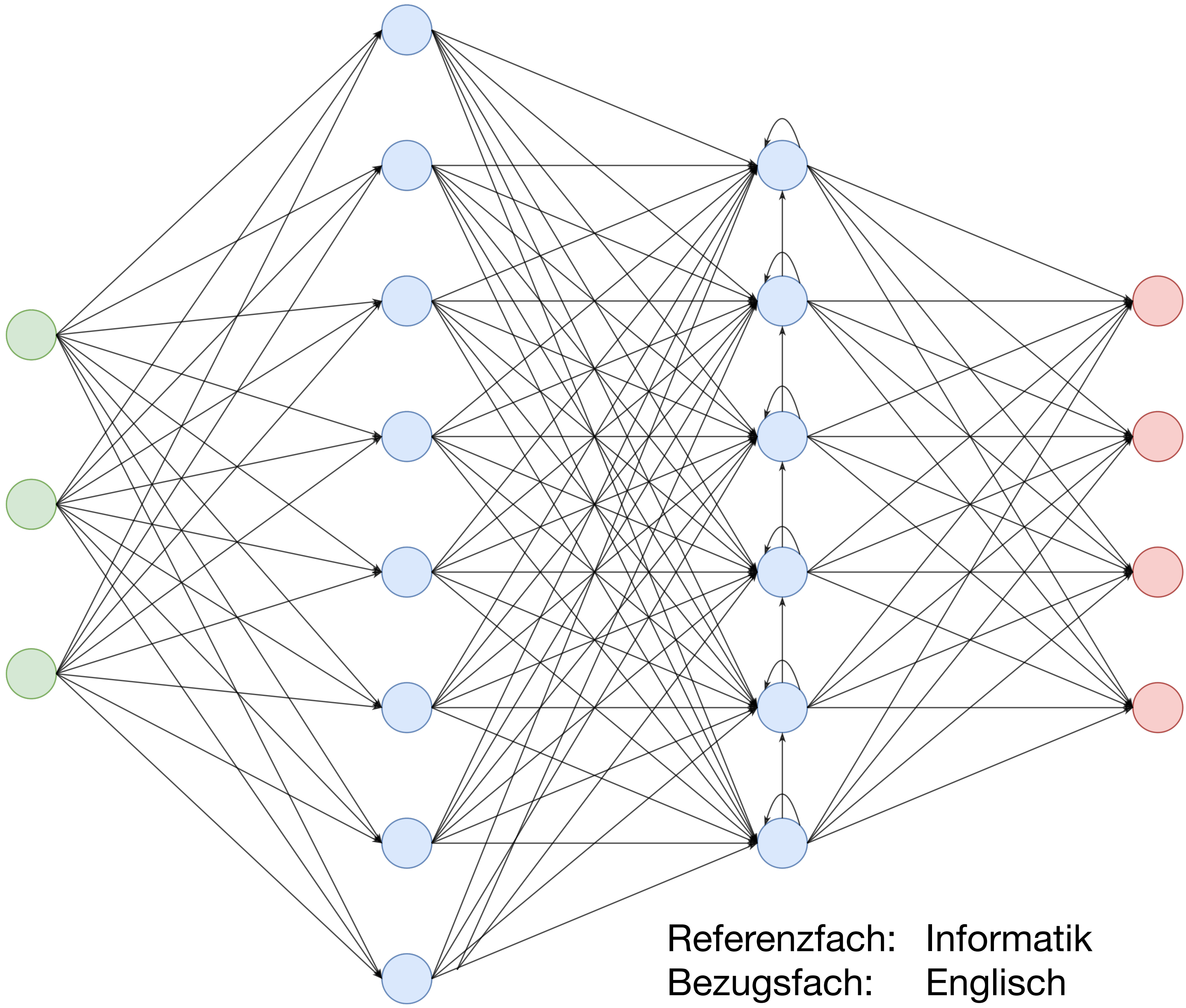
The opportunities presented by machine learning applied in education and the entailing challenges



Referenzfach: Informatik
Bezugsfach: Englisch

Oskar Thaeter

Prüfer: Jan Reher

Structure

Hypothesis

AN AI REVOLUTION
After the AI-Winter in 1990, created by unnaturally high expectations from end-users, exaggerated advertisement in the media and inflated assurances by developers, an “AI spring”, fuelled by field advancements and an ever-increasing computing power, described in Moore’s law[1], is upon us. While machine learning is not a solve-all solution, it promises to revolutionize industries and workforces. Taking over repetitive tasks from humans, machine learning not only presents opportunities, but also poses challenges of political, social and ethical nature.

AN EDUCATIONAL SYSTEM THAT CAN’T KEEP UP
There is a shortage of teachers in primary and secondary schools[2]. Meeting pupil demand while keeping the quality of current education is adding up to a “herculean task”[3].
The increasing number of pupils was previously underestimated[4].

DATA AS A RESOURCE
Every school day, every hour, pupils across the country write, summarize, analyse, outline, and read texts. Every year, over 400 thousand pupils[5] take extensive, state-centric exams in Germany, for multiple subjects. In the UK (except Scotland), over 700 thousand pupils[6] take exams in multiple subjects under one of five nation wide examination boards every year. This is an absurd amount of data and content not being used to anyones advantage besides making new exams. Marking over 5 million exams[7] is also no easy feat[8]. So all this data is either wasted or a burden to go through.

HYPOTHESIS
And this is exactly what machine learning is good at: processing a vast amount of data and doing repetitive tasks which display a certain pattern. So my proposal is: form datasets out of students essays, already marked exams and other written text to fuel research in the field of machine assisted education. Relieving teachers of marking could possibly save many hours per teacher. This gained time could be used by teachers to focus on the individual needs of their students. Therefore, **machine learning ought to improve education extensively**.

Machine learning

ARTIFICIAL INTELLIGENCE
In general terms, artificial intelligence (AI) is the attempt to recreate and or simulate human intelligence[9]. Therefore the goal of artificial intelligence is to achieve a general artificial intelligence. Because the term “intelligence” does not have a generally accepted, formal definition, the definition of “AI” changes over time and throughout the progress in the field and becomes more and more a philosophical rather than a purely technical question. Subfields of artificial intelligence are generally defined by approaches, goals or philosophical values.

WHAT IS MACHINE LEARNING?
The process of computers changing the way they carry out tasks by learning from new data, without a human being needing to give instructions in the form of a program.[10]
ML is a subfield of AI, using mathematical methods to derive a mathematical model, which shall solve a specific task with varying input but the same input structure. Generally, the broader the task, the more advanced the model has to be. In order to gain said functioning model it has to be “trained”. Training consists of attempting to solve the given task with the model over and over again, which adjusts its values, structure etc. Basically, training uses the principle of learning from past mistakes thus maximising the chance of future success.

ARTIFICIAL NEURAL NETWORKS

A neural network is a simple information processing algorithm[11].

Artificial neural networks are graphs consisting of neurones and connections between said neurones. These neurones are divided into groups, so called layers. ANNs begin with an input layer, where the input values are “fed” into the network, and end with an output layer, which consists of one ore more neurones holding a value between 0 and 1.

In between layers can have a range of functions, while a core layer, just consisting of simple neurones, is the most conventional type.

Other layers, among others, include: convolutional layers, pooling layers, recurrent layers, nominalization layers.[12]

The different types of layers are chosen for different types of ANNs, for example convolutional layers are used in convolutional neural networks (CNN) while recurrent layers are used in recurrent neural networks (RNN).

A neurone is made up of its input value, its output value and its activation function. Connections connect one neurone of one layer with one neurone of the next layer. These connections exist between every neurone of two subsequent layers, i.e. every neurone of the first layer is connected to every neurone of the second layer. Each connection is assigned a weight. A propagation function computes the input of a neurone by adding up every weighted output of the preceding neurones.

The general structure of ANNs is modelled after the neural network making up a human brain.

Application

THE PROBLEM

Our problem starts with most of the data being produced in schools being analogue, i.e. in handwriting rather than electronic signals. That makes it impossible to capitalize on the data in question. For this to change, a promising approach is using machine learning to translate the analogue data into machine readable digital data. To put it simply: Translate a picture of fixed size of a written word into according digital output.

THE DATASET

As a dataset we will use a novel approach to make training and modifying our model easier: a synthetic dataset which is of virtually infinite size and has near infinite possible variations. Other advantages include the data being of no cost, total control over the type of data and assurance of the data being correctly labeled and intact.

The aforementioned variations include image noise, position of the word, rotation of the word and most importantly the used font.

To simulate many different handwritings we use 23 distinct fonts in total. These 23 can be categorized into 5 font pools:

- “full” (all 23 fonts)
- “narrow” (in addition to Courier also includes more freeform fonts, which all use similar character forms)
- “block” (fonts strictly using block letters)
- “fancy” (cursive fonts with different structures for some characters)
- “False” (Courier, functions as a reference font)

This essentially gives us 5 different datasets, which we will compare in training to see how different writing styles effect training and end performance.

THE MODEL

Beginning a model with the supposed perfect architecture is rather rare, the best we can do is taking an educated guess as our starting point and adjusting the model architecture according to training performances of slightly changed layers and parameters.

Essentially trial and error is our best choice.

Because our problem needs us to translate a picture, a great starting point is a Convolutional Neural Network (CNN).

A CNN is actually just a neural network with a bunch of filters put in front of it. In praxis this means we will not have to take in every pixel of the picture, which would make our model much too large to train comfortably, but that we can break down the picture into many smaller groups of pixels and analyze these for features associated with distinct characters.

Now we have to ask ourselves what we want to detect in the picture. Would we like to just translate the displayed word or is it better to detect every character on its own to make up the word out of the detected characters.
Though the approaches might seem rather similar, they decide which kind of model we will be using. Also the application has two distinct names in machine learning, image classification and optical character recognition. Image classification requires a category or class for each word and typically uses a CNN to perform its task. Optical character classification does not detect a whole word but rather detects the characters making up the word. In practice that means that recurrent neural networks (RNN) are used to output a sequence instead of a class. In turn that means that we have the characters of the alphabet as our classes.

In conclusion, we will use a convolutional neural network to detect features in the input picture and the features then will be fed into a recurrent neural network to discern distinct characters. Our output will be a sequence of characters, which will be achieved by decoding the output of the RNN using a connectionist temporal classification (CTC) function.

BUILDING THE MODEL
To train a model we first have to build a trainable model. For that we are using Keras, a high-level neural networks API[13], with TensorFlow as the backend. Because training a neural network requires a large amount of computations it is beneficial to use a GPU to train the model. Google Colaboratory provides a GPU for research purposes to anyone for free, while also running entirely in the cloud, requiring no setup and being accessible via browser. In order to not make the model building unnecessarily difficult we will use an example from the Keras website[14] as a starting point, allowing us jump a few obstacles in advance. The part that interest us is primarily the model itself, the example using 2 convolutional layers, each followed by a pooling layer, a conventional neurone layer before two bidirectional GRU layers (GRU is a type of recurrent layer) and a final activation layer. Note that though the CTC function comes after the activation layer, it is not part of the model itself.

This is indeed a good starting point, as it already follows the general structure of our imagined model. The first we will change is the recurrent stack. Because Gated Recurrent Units (GRU) are mostly used because of their computational efficiency we will instead use Long Short Term Memory layers (LSTM). Since both layers aim to do the same, tackling the problem of vanishing and exploding gradients, and it isn't entirely clear which is better[15] we will move forward with the LSTM which has been longer around, first proposed in 1997[16], rather than with the relatively new GRU[17], first introduced in 2014.

Because we are using a CTC function, we have to use the functional API[18] of Keras, which allows us more freedom in connecting layers with each other and most importantly allows for multiple inputs to one layer.

TRAINING THE MODEL

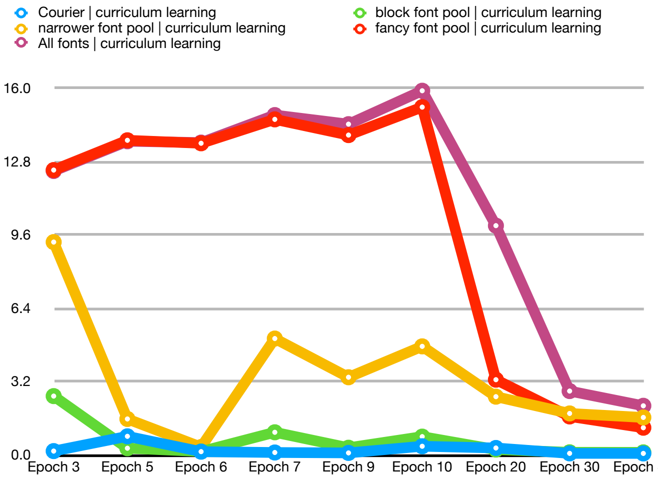
To train our model we have to apply a gradient descent optimiser to it. There are numerous optimisers out there but not all work equally well on different model architectures. The simplest optimiser is stochastic gradient descent (SGD), with more advanced algorithms being for example RMSProb and Adadelta. It is often worthwhile to empirically find out which optimiser works best in the current situation.

In order to be able to train our model we also need a loss function. In our case this function is the CTC function. Loss is a number indicating how bad the model's prediction was[19]. The optimisers will try to change the weights of the model in a way that will minimize the loss and in that way maximize the accuracy of our model.

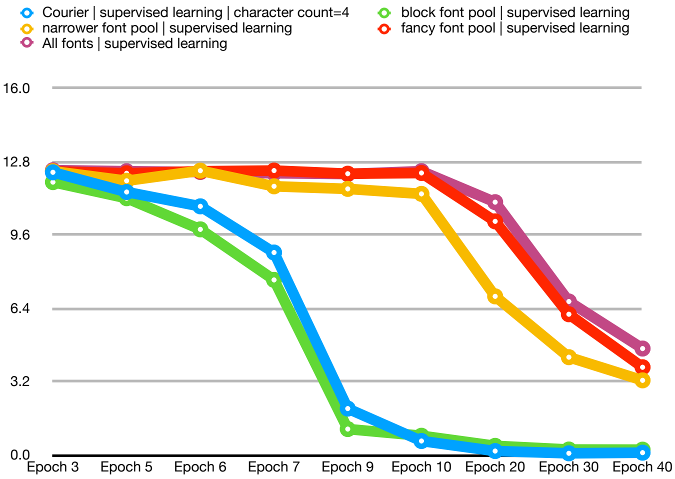
Training is divided into multiple epochs, which are also again divided into multiple batches.
An epoch has a training part and a validation part. In the training part the model is optimised, in the validation part the model is confronted with never before seen input so that its effectiveness can be assessed. In our case the model will “see” 16000 images,
80% (12800) of which will be used for training with the remaining 20% (3200) being used for validation. The 12800 training images will be delivered in 400 batches of 32 images each. In addition to the validation step at the end of every epoch the model will be tested on
256 fresh images. With these the mean edit distance and the mean normalized edit distance is determined. Edit distance is a measurement used to determine how similar or dissimilar two strings are, i.e. how close the decoded output to the input label is.

As an indication of how the model is improving over the course of the training it is best to take a look at the validation loss of successive epochs. The time required for training a model is effected by many factors, although the most influential ones are the number of parameters the model has, the number of images per epoch and the hardware used for training. Because we are using a GPU (NVIDIA K80 Tesla) to train our model, training is much accelerated compared to using a CPU. For example: an Epoch of 16000 images takes a model with over 12,000,000 parameters around 260s to go through, the same Epoch takes a model with 650,000 parameters only up to 50 s.

We will try out two different training techniques. First, the conventional learning technique, where the dataset will not be changed during training, and then curriculum learning, where the variations of the data are gradually increased over the first 10 epochs. We will look at every font pool’s performance to assure that any findings may apply to all datasets.



curriculum learning

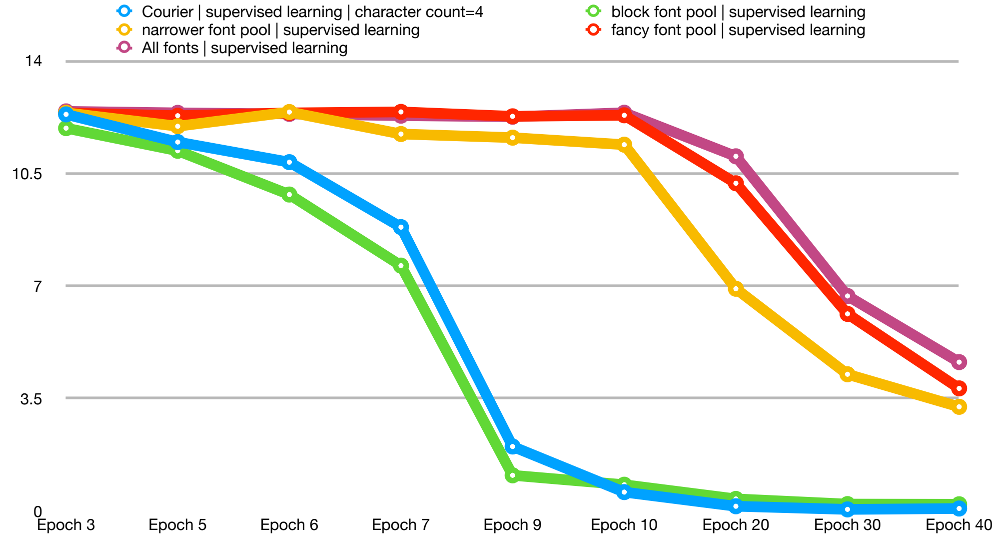


conventional supervised learning

Comparing the two graphs, we can make out that, even though the models using curriculum learning start off worse for the elaborate fonts, in the end they all achieve a lower validation loss.
And because the actual improvement of our models over the epochs do not matter in the end as long as the accuracy/validation loss is agreeable, moving forward we will use curriculum learning to train our models. Our findings are also supported by the paper which originally introduced curriculum learning to train deep neural networks. “we note that the observed advantage of CL is more significant when the task is more difficult”

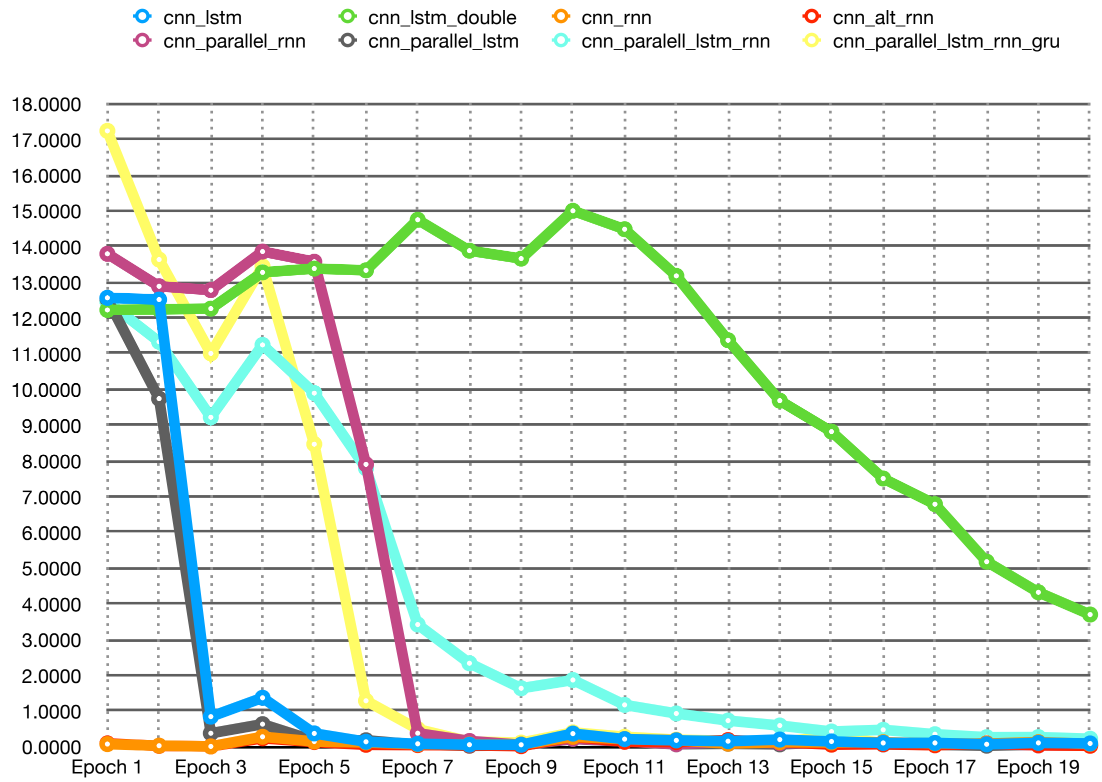
ANALYSING AND IMPROVING THE MODEL

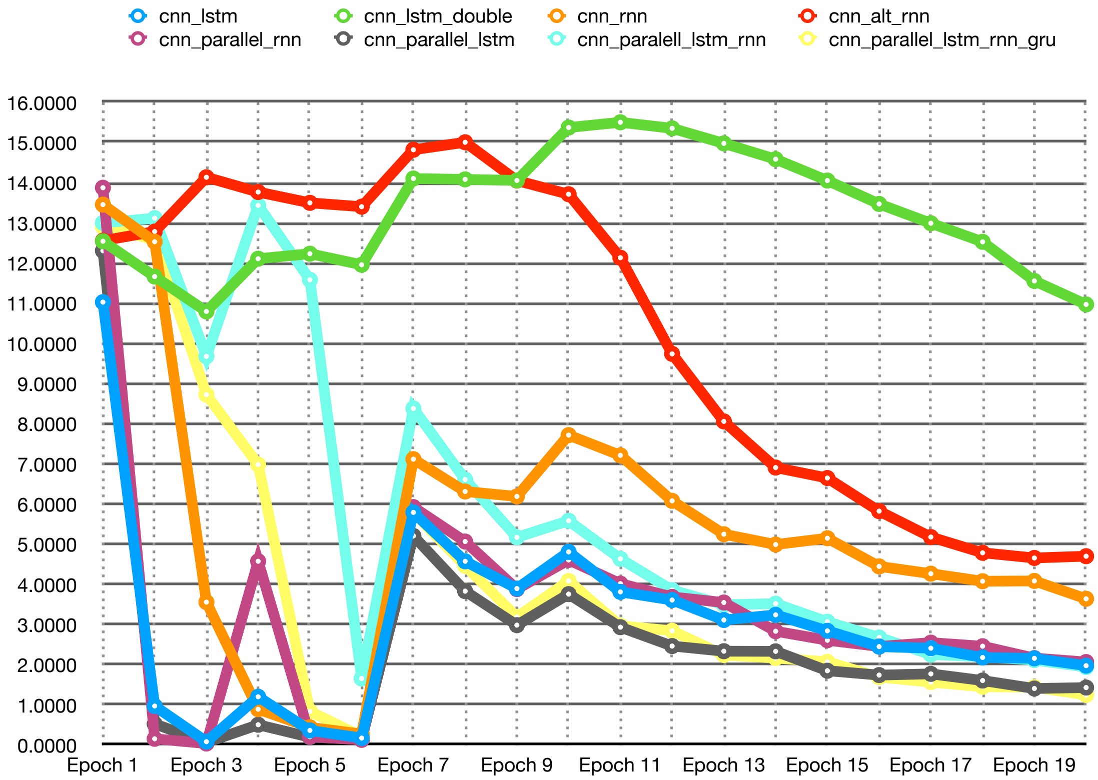
Improving the model goes hand in hand with training the model, because improvement is mostly achieved by trail and error and making educated guesses. That means we will take an empirical approach to improve our model.
First, we will asses the dataset: How complex is the task we are tackling? To answer the posed question we compare the training validation loss of our original model on the aforementioned font pools:



As we can see, the graph can be grouped into two sections. One section made up of “easier” datasets to learn and the other made up of more “difficult” datasets. This behaviour is simply explained by the structure of the data. The second section of the graph includes the datasets which are made up of many fonts, having many variations between them. While the remaining two datasets only consist of block letter fonts, with the Courier dataset even only consisting of a single font. In conclusion, we observe a definite spike in “difficulty” going from block fonts to cursive fonts. The datasets where all used on the same model, a CNN connected to a LSTM.

Now we may try out different layers and layer architectures. To asses their different strengths and weaknesses we will be training the alternative models on the Courier dataset and the full dataset.
Evaluating the models after training them on
two different datasets for 20 epochs each
we reach the following conclusions:
- more parameters do not necessarily mean better results
- innovative designs can pay off
- complexion does aid performance





THE FINAL MODEL

Even the best performing model “cnn_parallel_lstm_rnn”, after training for 60 epochs, with a final validation loss of 0.5284 and a total of 906 correctly predicted words out of 1000, is no way near good enough to deploy in the real world. Tested on real data, i.e. scanned hand-written words, the model does not correctly predict a single word out of 10 examples.

Does this mean this attempt at a machine learning OCR model failed?
I would argue: No, it did not. I have demonstrated common techniques and methods used in machine learning like synthetic datasets, curriculum learning, convolutional and recurrent layers, comparing network architectures and more importantly shown, that recognizing handwritten cursive words is difficult for machine learning. My main takeaway from this project is, that although it isn’t quite there yet, OCR will become better at recognising cursive words and thus enable using analogue data as datasets for future endeavours.

Opportunities

DATA DEMOCRATISATION

In order to make advances in machine learning and AI accessible to anyone who could benefit from them, one approach is open source. Widely used in traditional fields of computer science, open source is the idea of making source code readily available to anyone, while it can even go further, open source projects often becoming a collaborative effort of hundreds of people all working on the same code via platforms like GitHub. Open AI is the most prominent open source AI company, recently pioneering advances in language models with it’s GPT-2 model[20]. The company commits itself to make data of new research available to the public[21].
"arXiv.org", an open access repository for preprint scientific papers hosted and owned by Cornell University, is an example for sharing scientific advances, making innovations and new findings at no cost available to anyone.

While data democratisation helps business people make better decisions[22], sharing the data that is created in the school and educational system could help people understand the shortcomings of the current educational standards and therefore help them make informed decisions regarding the educational system.
OCR and HTR have great potential to work across borders and language barriers, with possibly only minor adjustments needed to adapt to a new language. A greater challenge is enumerating a new dataset for every new alphabet. And this is where open source should be employed, because datasets are most often the most expensive part of a machine learning model.

CURRENT TECHNOLOGY
While there are datasets for automated scoring, for example the Automated Students Assessment Prize dataset consisting of 12,978 scored essays with a total of 2,889,684 words[23], the size threshold to train a sufficiently reliable scorer, that not only assess grammar and orthography but also can accurately evaluate the content coherence of an essay according to specified exam expectations, is much higher with current technology.
So how could this problem be solved? The best way to enumerate enough data is by directly going to the source: schools. With an advanced model based on the proposed OCR model it would be possible to reliably translate analogue text into digital data. This would mean that a dataset could be enumerated consisting of original exams, pupil's work and the marked exams. With this dataset a machine learning model could be trained on correctly marking exams based on the exam and the pupil's work.

Because such a solution is still far from being realised, an open source dataset would spur further scientific research and help development in this direction.
Presently, automated essay scoring (AES) systems are either feature-engineered models, i.e. number of words, grammatical errors etc., or end-to-end models driven by deep learning, these being able to extract semantic features. A recent paper by Liu et al. (2019) proposes combining the two approaches[24]. Dong et al.(2016) employs a convolutional neural network to automatically learn features[25].

Challenges

TECHNOCHAUVINISM
The idea, that any problem can be solved with just the right amount of technology is described by Meredith Broussard in her book "Artificial Unintelligence: How Computer Misunderstand the World"[26] as "technochauvinism". Broussard defines the term as: "an unwavering faith that if the world just used more computers, and used them properly, social problems would disappear and we'd create a digitally enabled utopia."[27]

Considering the approach we have taken until now, we are certainly guilty of overlooking unquestionable obstacles like ensuring equal access to resources and technologies.
How are schools supposed to make use of a theoretical automatic scorer, if they can't even afford[28] to buy enough books for their students[29] not to mention a scanner to digitize sheets of paper[30]. Unequal access to digital infrastructure is part of the reason why we have to translate analogue writing into digital data in the first place, most schools not being able to provide the electronic devices needed to conduct an exam completely paperless. So no matter how sophisticated our scorer model will be, making it accessible to everyone is another problem.

THE NEED FOR HUMAN EXPERIENCE
Marking complex exams like A-Level english papers can require more than just comparing scripts with the mark scheme, as exhibited by the "A-Level English Language Paper 1 Mark Scheme" explicitly stating: "It is important to be open minded and positive when marking scripts."[31]
This is an inherently human approach that is required of the marker. This is a clear reprove of the current abilities of narrow artificial intelligence, evidently demanding consciousness.

BIAS

Les Perelman, of the Massachusetts Institute of Technology, rebukes in his 2014 research article "When 'the state of the art' is counting words"[32] the claims of the article "State-of-the-art automated essay scoring: Competition results and future directions from a United States demonstration", which stated: "Automated essay scoring appears to have developed to the point where it can consistently replicate the resolved scores of human raters in high-stakes assessment." Perelman argues, that "The state-of-the-art referred to in the title of the article is, largely, simply counting words."[33], hinting at the most significant criticism of the current state of automated essay scoring: favouring complex vocabulary and text length over content and coherence. Biases can also stem from failures in the development process like selecting a dataset with inherent biases against groups of people or a not diverse enough development team[34].

Conclusion

DOES MACHINE LEARNING OFFER A VIABLE SOLUTION TO EDUCATIONAL ISSUES?

As we explored the idea to use exams and scripts to train an algorithm to mark them, we have come across opportunities and advantages this approach presents but also challenges to it. While the use of artificial intelligence in every possible field is inevitable, the use of it in education is still far off, because implementing it will first pose seemingly insurmountable obstacles, e.g. legal, cultural, technical issues, before AI establishes itself as an integral part of the educational system. While the use of AI might seem like a viable solution to some problems, it should not be applied to every problem. Could artificial intelligence relieve teachers of marking?

Yes, but is marking really the biggest problem in education today? American pupils cite[35] pressure to succeed, not engaging enough lessons and an emphasis on grades as having a negative impact on their education. They also suggest more resources and support for teachers to improve American education. While issues vary across borders, teachers not having enough time for marking papers does not necessitate a big technological undertaking, but only the educational system lowering the demands on a single teacher.

Considering the previously presented facts and taking them into account, while I do not think the proposed use of machine learning/AI is a viable solution to the issues facing education today, schools and classrooms powered by AI are unavoidably in the future. Artificial intelligence presents a way to fight symptoms, but it is not a solution to the root causes of societal issues.

Literature and Sources

[1] Moore, Gordon E. (1965): "Cramming more components onto integrated circuits"
(Electronics, Volume 38, Number 8, April 19, 1965)
URL: https://newsroom.intel.com/wp-content/uploads/sites/11/2018/05/moores-law-electronics.pdf
[Retrieved 09/12/2019]

[2] Klovert, Heike; Laurenz, Nike und Quecke, Franca (2019): "Warum in Deutschland so viele Lehrer fehlen: Grund- und Berufsschulen spüren es besonders stark: Bundesweit fehlen Lehrer. Woran liegt das? Wie steuern die Länder gegen? Die wichtigsten Antworten im Überblick."
URL: https://www.spiegel.de/lebenundlernen/schule/lehrer-mangel-in-deutschland-warum-so-viele-paedagogen-fehlen-und-wo-a-1280209.html
[Retrieved 09/12/2019]
article in "Spiegel Online"

[3] Zorn, Dr. Dirk (2019): "Lehrermangel in Grundschulen bis 2030 größer als bislang erwartet"
URL: https://www.bertelsmann-stiftung.de/de/themen/aktuelle-meldungen/2019/september/lehrermangel-in-grundschulen-bis-2030-groesser-als-bislang-erwartet/
[Retrieved 09/12/2019]
"Bertelsmann Stiftung" study
"Die Bewältigung des Lehrermangels ist eine Herkulesaufgabe. Die Zahl der Schülerinnen und Schüler wächst dynamischer als angenommen."

[4] ibid.
"Auch 2030 werden die Schülerzahlen absehbar höher liegen als von der KMK erwartet"

[5] Statistisches Bundesamt (2018): "Absolventen/Abgänger nach Abschlussart und Geschlecht"
URL: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Schulen/Tabellen/liste-absolventen-abgaenger-abschlussart.html;jsessionid=36A47054D1EEB35C7D3C66BC068CF4D0.internet711
[Retrieved 03/01/2020]

6 BBC News (22.08.2019): "GCSEs 2019: Where are we at with new grades 9-1?"
URL: https://www.bbc.com/news/education-48993830
[Retrieved 03/01/2020]

[7] The Office of Qualifications and Examinations Regulation (2018): "Entries for GCSE, AS and A level: Summer 2018 exam series"
URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/712450/Report_-_summer_2018_exam_entries_GCSEs_Level_1_2_AS_and_A_levels.pdf
[Retrieved 03/01/2020]
annually published, official statistics

[8] AQA awarding body: "How Exams Work"
URL: https://www.aqa.org.uk/about-us/what-we-do/getting-the-right-result/how-exams-work
[Retrieved 03/01/2020]

[9] Kelleher. John D.: Deep Learning
Cambridge, Massachusetts: The MIT Press, 2019
page 251, ll. 7-10

[10] Cambridge Dictionary: "machine learning"
URL: https://dictionary.cambridge.org/dictionary/english/machine-learning
[Retrieved 08/12/2020]

[11] Kelleher. John D.: Deep Learning
Cambridge, Massachusetts: The MIT Press, 2019
page 254, ll. 5-6

[12] Keras Documentation: "About Keras layers"
URL: https://keras.io/layers/about-keras-layers/
[Retrieved 08/12/2020]

[13] Keras Documentation: “Keras: The Python Deep Learning library”
URL: https://keras.io/#you-have-just-found-keras
[Retrieved 01/02/2020]

[14] Keras Documentation: “Optical character recognition”
URL: https://keras.io/examples/image_ocr/
[Retrieved 01/02/2020]
exemplary python code using Keras

[15] Chung et al. (2014): “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling”
page 7, scientific publication
URL: https://arxiv.org/pdf/1412.3555v1.pdf
[Retrieved 01/02/2020]
preprint scientific paper comparing different types of recurrent units in RNN

[16] Hochreiter, Sepp und Schmidhuber, Jürgen: “Long Short-Term Memory”
(Neural Computation, 9(8):1735–1780, 1997)
scientific publication
URL: https://www.bioinf.jku.at/publications/older/2604.pdf
[Retrieved 06/02/2020]
scientific paper introducing Long Short-Term Memory

[17] Cho et al. (2014): “Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation”
scientific publication
URL: https://arxiv.org/abs/1406.1078
[Retrieved 01/02/2020]

[18] Keras Documentation: “Guide to the Functional API”
URL: https://keras.io/getting-started/functional-api-guide/
[Retrieved 01/02/2020]

[19] Google Machine Learning Crash Course: “Descending into ML: Training and Loss”
URL: https://developers.google.com/machine-learning/crash-course/descending-into-ml/training-and-loss
[Retrieved 05/02/2020]
part of an online course on machine learning by Google

[20] Radford et al. (2019): “Language Models are Unsupervised Multitask Learners”
URL: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf
[Retrieved 03/01/2020]
scientific publication

[21] OpenAI (2018): “OpenAI Charter”
URL: https://openai.com/charter/
[Retrieved 03/01/2020]

[22] Marr, Bernard (2017): “What Is Data Democratization? A Super Simple Explanation And The Key Pros And Cons”
(Forbes; 24.07.2017)
URL: https://www.forbes.com/sites/bernardmarr/2017/07/24/what-is-data-democratization-a-super-simple-explanation-and-the-key-pros-and-cons/#716c97fb6013
[Retrieved 03/01/2020]

[23] The Hewlett Foundation (2012): “The Hewlett Foundation: Automated Essay Scoring”
URL: https://www.kaggle.com/c/asap-aes/data
[Retrieved 03/01/2020]
datasource for ASAP AES competition, this dataset is most often used in automated essay scoring systems

[24] Jiawei Liu, Yang Xu, und Yaguang Zhu (2019): “Automated Essay Scoring based on Two-Stage Learning”
page 1, scientific publication
URL: https://arxiv.org/pdf/1901.07744.pdf
[Retrieved 04/02/2020]
preprint scientific paper proposing a Two-Stage Learning Framework, using feature-engineered and end-to-end AES methods

25 Fei Dong and Yue Zhang: “Automatic Features for Essay Scoring – An Empirical Study”
(Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pages 1072–1077, Austin, Texas, November 1-5, 2016.© 2016 Association for Computational Linguistics; 2016)
URL: https://www.aclweb.org/anthology/D16-1115.pdf
[Retrieved 04/02/2020]
scientific paper on automatic feature learning in automatic essay scoring

26 Broussard, Meredith: Artificial Unintelligence: How Computer Misunderstand the World
Cambridge, Massachusetts: The MIT Press, 2018

27 ibid. page 8, ll. 10-13

28 ibid. page 60, ll. 6-7
“there’s simply too little money in the education budget”
note that this and the following quotes concern the Philadelphia school system

29 ibid. page 60, ll. 7-9
“The [required] textbook costs $114.75. However, for 2012-2013, [the school] was only allocated $30.30 per student to buy books [for ever subject]”

30 ibid. page 57, ll. 5-7
“Philadelphia teachers spend an average of $300 to $1,000 of their own money each year to supplement their $100 annual budget for classroom supplies”

31 A-Level English Language 7702/1 Language, The Individual and Society (New) Final Mark Scheme June 2017
page 3
“It is important to be open minded and positive when marking scripts.”
URL: https://filestore.aqa.org.uk/sample-papers-and-mark-schemes/2017/june/AQA-77021-W-MS-JUN17.PDF
[Retrieved 04/02/2020]

32 Les Perelman: “When “the state of the art” is counting words”
(Assessing Writing 21 (2014) 104–111; 2014)
URL: http://lesperelman.com/wp-content/uploads/2015/09/Perelman-State-of-the-Art-is-Counting-Words.pdf
[Retrieved 04/02/2020]
research article on the current state of automated essay scoring

33 ibid. page 104, Abstract

34 Broussard, Meredith: Artificial Unintelligence: How Computer Misunderstand the World
Cambridge, Massachusetts: The MIT Press, 2018
page 67, l. 8
“Computer systems are proxies for the people who made them.”

35 The Learning Network: “What Students Are Saying About How to Improve American Education: An international exam shows that American 15-year-olds are stagnant in reading and math. Teenagers told us what’s working and what’s not in the American education system.”
(The New York Times; 19.12.2019)
URL: https://www.nytimes.com/2019/12/19/learning/what-students-are-saying-about-how-to-improve-american-education.html
[Retrieved 04/02/2020]
The New York Times article about the results of a prompt asking students how American education could be improved in light of stagnant performance of American teenagers in reading and math according to a study by the Program for International Student Assessment