Списание „Математика, компютърни науки и образование“ Том 4, кн. 1, 2021
Journal of Mathematics, Computer Science and Education Vol. 4, No. 1, 2021

DOI: 10.54664/GCHA4027

On Some Aspects of Modern AI Applications

Hristiyan Dimitrov

Abstract: *This paper presents modern AI applications using real life company projects, which try to solve problems related to the self-driving car industry, medicine and the overall quality of life. By utilizing big data, efficient algorithms and proper modelling, these projects make exceptional progress in their respective fields. The paper examines the underlying foundations of the projects and the challenges they face.*

Keywords: *artificial intelligence (AI), machine learning, artificial neural networks, diabetic retinopathy (DR), amyotrophic lateral sclerosis (ALS), recurrent neural network (RNN), RNN Transducer (RNN-T), self-driving cars, automotive industry, automatic speech recognition (ASR), generative adversarial networks (GANs), Listen, Attend and Spell (LAS), function rating scale (FRS).*

Introduction

In our modern era, we are exploring the capabilities of AI and all of its subcategories. Deep learning has made exceptional progress in unstructured data where even we as humans do not see the clear patterns, whereas computers do. There are many cases in which we know the problem, but just vaguely see the solution. Here we have explored those concepts in real life projects where AI helps us see where our focus needs to shift. The projects in this paper show the importance of bringing all people to the modern world, including those with specific disabilities and lower income.

Deep Learning for Detecting Diabetic Retinopathy

According to the NHS (National Health Service), diabetic retinopathy (DR) is a complication of diabetes caused by high blood sugar levels damaging the back of the eye (retina). It can cause blindness if left undiagnosed and untreated. It is also one of the leading causes for blindness worldwide, approximating 95 million patients diagnosed with diabetes. Retinal photography is a widely accepted screening tool for DR. However, the lack of ophthalmologists worldwide, especially in poorer countries like India and continents like Africa, can cause an issue when identifying and treating this disease. Researchers at Google addressed this issue back in 2016 [2, 3] by deciding to create a model that can accurately predict the severity of DR in patients diagnosed with diabetes. Their first approach – machine learning, did not prove effective in terms of covering a wide variety of classification tasks. The deep learning technique proved to be better at finding explicit features just by looking at the pictures and using an optimization algorithm called *back-propagation* to indicate how a machine should change its internal parameters to best predict the desired output of an image (Figures 1 and 2).

The project underwent several stages of acquiring data from hospitals in the US and India. After that, a chunk of the images was graded by professional ophthalmologists for the presence of diabetic retinopathy, diabetic macular edema, and image quality. The International Clinical Diabetic Retinopathy Scale was used to identify how severe the DR is – none, mild, moderate, severe or proliferative. The process of training a neural network to perform a given task is called *deep learning*. The large dataset which Google had accumulated and which had been graded by professionals, laid a solid foundation for a “well-trained” model. The development was divided into two parts – the first one for training (80%) and the second one for optimization (20%). After rigorous testing, a single network was trained to make multiple binary predictions for the severity of DR.

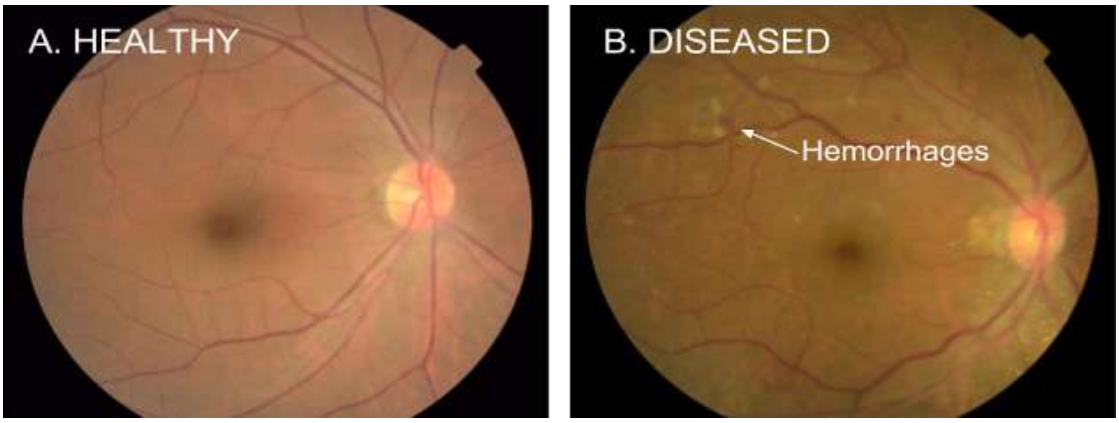

Figure 1. Examples of retinal fundus photographs taken to screen for DR. The image on the left is of a healthy retina, whereas the image on the right is of a retina with referable diabetic retinopathy due to a number of hemorrhages (red spots) [2]

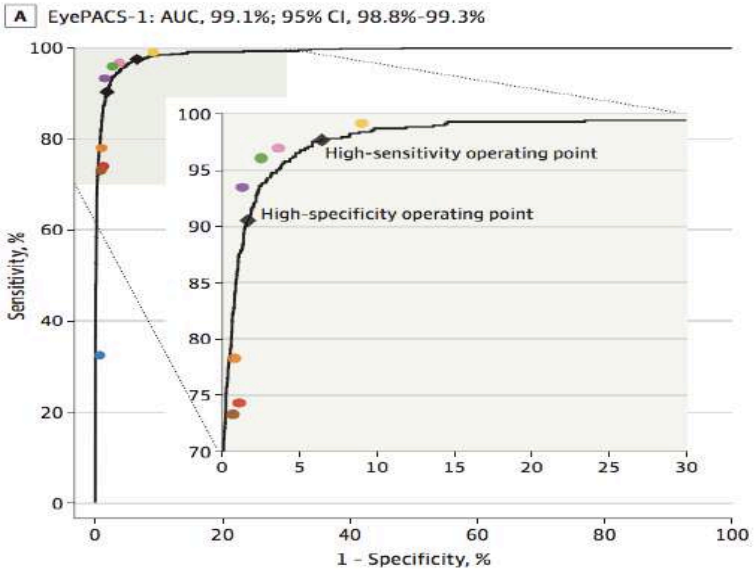

Figure 2. Performance of the algorithm (black curve) and eight ophthalmologists (coloured dots) for the presence of referable diabetic retinopathy (moderate or worse diabetic retinopathy or referable diabetic macular edema) on a validation set consisting of 9963 images. The black diamonds on the graph correspond to the sensitivity and specificity of the algorithm at the high sensitivity and high specificity operating points [2]

Project Euphonia

Automatic speech recognition (ASR), also known as speech recognition in computer science, is a technology that allows the conversion of human language into a computer string in most cases. Most virtual assistants like Siri, Alexa, etc., use this technology to understand us and respond in a proper manner. However, those assistants have been tuned to work with the general public. A team at Google identified an issue [4, 5] when ASR could not understand speech-impaired individuals. People with amyotrophic lateral sclerosis (ALS), for example, are greatly impacted by one of the common symptoms – slurred speech. Thus, Project Euphonia was created to help adjust each model to those affected by different severities of slurred speech.

The Project Euphonia team faced two challenges: individuals speaking in different ways, and the lack of data when adjusting the models to every individual. The first of the architectural approaches is the RNN Transducer (RNN-T) which is a neural network architecture consisting of encoder and decoder networks. The second approach – Listen, Attend and Spell (LAS), is an attention-based, sequence-to-sequence model that maps sequences of acoustic properties to sequences of languages. The focus was on the bidirectional RNN-T with some tests run on the LAS as well. The datasets that were used in training the models came from Google's partnership with the ALS Therapy Development Institute and the L2-ARCTIC dataset of non-native speech (Figure 3).

For the RNN-T, the team started by fine-tuning layers 1, 2 and 3 in fixed combinations (treating the decoder as a single layer) on both datasets, adjusting hyperparameters as necessary. For the LAS architecture, they fine-tuned various layer combinations and consistently found out that the best results from this network came from fine-tuning all layers. To better understand their models, the team looked at the patterns of phoneme errors when using ASR applied in ASL speech, and established that ASR errors in ALS speech were far more similar to regular speech errors after Euphonia fine-tuning.

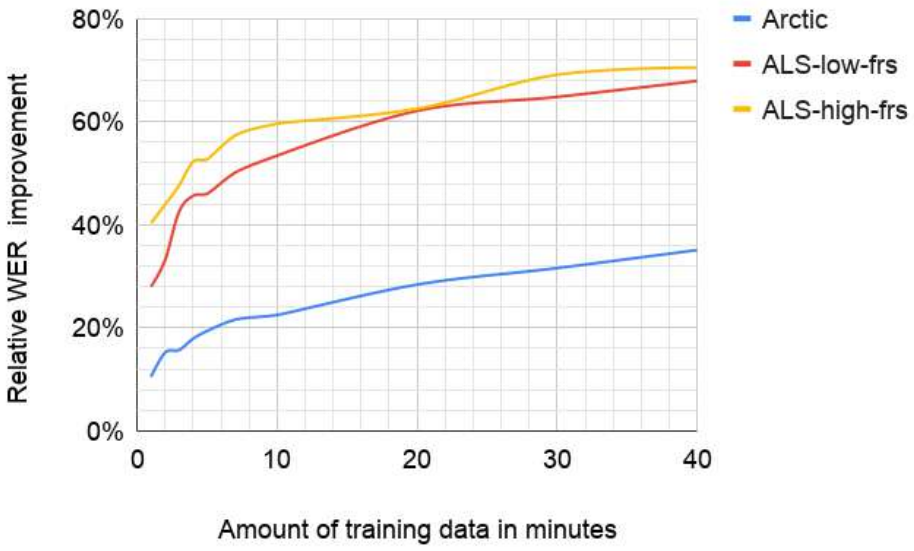

Figure 3. Low FRS corresponds to ALS speakers with low intelligibility (FRS 2, 1), while high FRS corresponds to ALS speakers with less severely impacted speech (FRS 3) [4]

Comma AI

Comma AI has been dealing with a problem concerning autonomous driving experience without having to buy a car, specifically designed to support that feature. Contrary to industry standards, Comma AI has been trying to sell a physical product that can be attached to supported car models. The datasets were recordings using mid-tier cameras mounted on the windshields of cars. The released video frames were 160 × 320 pixel regions from the middle of the captured screen.

Example data were also extracted from the sensors of the car, such as speed, steering angle, GPS, gyroscope, IMU, etc. The team's approach was based on an agent that learns how to clone a driver's behaviour and use these data to make simulation and predict the future. In the paper from 2016 [1], the architecture (Figure 4) used an autoencoder for dimensionality reduction and an action-conditioned RNN for learning the transitions. The first stages of development consisted of experimenting with different approaches for the autoencoders and cost functions. The idea of the team was to make the representation compact, which turned out to be almost 16 times smaller than the original data dimensionality, but preserving the road texture. A good autoencoder trained with generative adversarial networks (GANs) was

obtained, and the second stage of development was started where the transition model would be trained using a recurrent neural network (RNN) in the embedded space.

The learned transition model would keep the road structure even after 100 frames (Figure 5). *One of the challenges facing the team was related to simulating curves*, but events like passing lanes, approaching leading cars, etc. gave hope. The conclusion drawn from this paper is that instead of learning everything end-to-end, the first trained autoencoder with generative adversarial network-based cost functions generated realistically looking images of the road, which was then followed by the training of the RNN model in the embedded space.

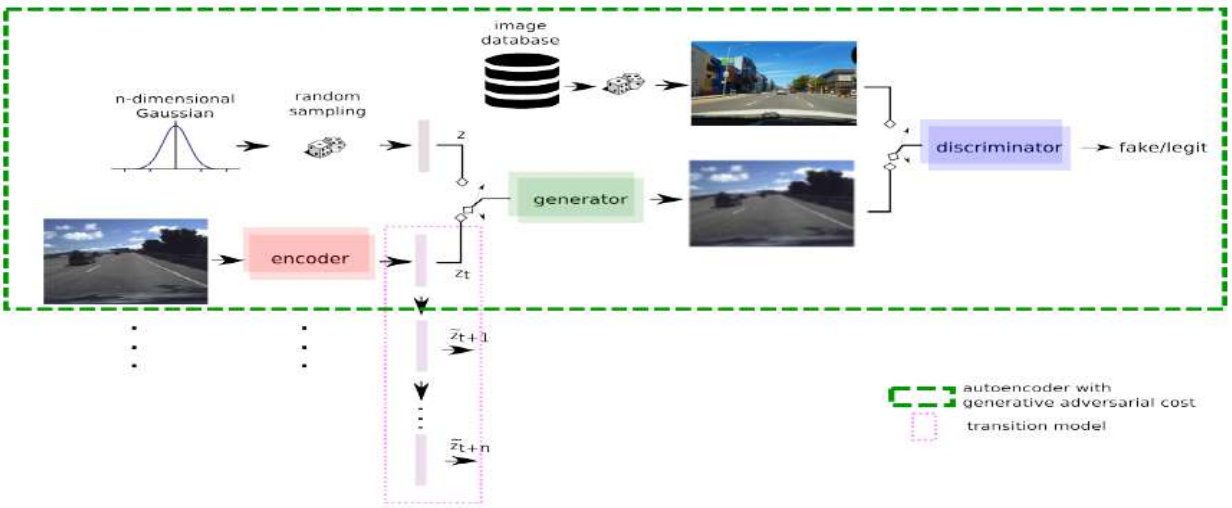

Figure 4. Driving simulator model: an autoencoder trained with generative adversarial costs coupled with a

Figure 5. Samples using similar fully convolutional autoencoders. Odd columns show decoded images, even columns show target images. Models were trained using generative adversarial networks cost function [1].

Conclusions

The common between the three projects is that well-trained models can be created with limited data or such that require professional expertise. They provide very insightful research into domains that

we are yet to explore and understand. The future of these projects can be expanded into some very serious research papers, broadening horizons for new problems and new, better solutions.

REFERENCES

[1.] **George Hotz, Eder Santana.** Learning a Driving Simulator [3 August 2016]

[2.] **Lily Peng, Varun Gulshan.** Google's blog about their Deep Learning for Detection of Diabetic Eye Disease [29 November 2016] – Google AI Blog: Deep Learning for Detection of Diabetic Eye Disease

[3.] **Lily Peng, Varun Gulshan.** Paper on developing and validating the deep learning algorithm [13 December 2016] – Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs

[4.] **Joel Shor, Dotan Emanuel.** Google's blog about Project Euphonia [13 August 2019] – Google AI Blog: Project Euphonia's Personalized Speech Recognition for Non-Standard Speech

[5.] **Joel Shor, Dotan Emanuel.** Personalizing ASR for Dysarthric and Accented Speech with Limited Data [31 July 2019]

ABOUT THE AUTHOR

Hristiyan Dimitrov, Year 3 undergraduate in Software Engineering,
Faculty of Mathematics and Informatics, Department of Information Technology,
St. Cyril and St. Methodius University of Veliko Tarnovo, Bulgaria, Tel.: +359 882 240 073
Email: hristian893@gmail.com