Data Science Quantifying the World DS7333.4043

Lance Dacy
Reannan McDaniel
Shawn Jung
Jonathan Tan

Unit 12 Case Study

July 28, 2020

Table of Contents

Introduction	1
Replica Neural Network	2
Recommendations	2
Contrasting Practices	2

Introduction

In June of 2014, 3 professors from the University of Southern California-Irvine (D. Baldi, P. Sadowski, & D. Whiteson) were set on a mission to demonstrate that deep learning approaches can improve the power of collider searches for exotic particles.

According to James Trefil in his writing “The Dark Side of the Universe”, there are a number of theories that propose to explain what happened in the very early universe (the first small fraction of a second) that link elementary particle physics and cosmology to predict the existence of exotic particles that have yet to be discovered.

If these particles exist, they could contribute significantly to the dark matter in the universe. One category of exotic particles includes the WIMPs: Weakly Interacting Massive Particles, including the photino, the axion, and the squark. Because these particles are relatively massive, they would travel at relatively low velocities and therefore could be considered "cold". These particles are marked as "cold dark matter" (as opposed to "hot dark matter" like neutrinos).

Each of these particles could be very common and very hard to observe. The scientist’s purpose is to allow the unification of theories that govern physics in the first 10-35 seconds of existence. There is no evidence that such particles exist, but there is a quest is to prove that maybe they do.

According to the July 2014 paper’s abstract, collisions at high-energy particle colliders are a traditionally fruitful source of exotic particle discoveries. Finding these rare particles requires solving difficult signal-versus-background classification problems, hence machine-learning approaches are often used. Standard approaches have relied on ‘shallow’ machine-learning models that have a limited capacity to learn complex nonlinear functions of the inputs, and rely on a painstaking search through manually constructed nonlinear features.

Given that technology changes radically over a few months, let alone a few years, these 3 scientists decided to resurrect the original idea of deep learning with new tools to explore. Our goal in this case study is to apply technology of 2020 to see what advances have been made in 6 years with learning technology and how this might change the course of the research.

Replica Neural Network

Using Tensor Flow and the original 2014 paper's architecture, the goal was to recreate the model to evaluate current technology that could enable a more usable model than before. The model can be found in here.

Recommendations

Any recommendations or modifications you'd make to the approach taken by the researchers (i.e., state any proposed changes along with the expected improvements or impact)

Of the many recommendations using today's technology, one of the first ones would be the use ReLU (Rectified Linear Unit) instead of Tanh as an activation function. The largest advantage of ReLu is a non-saturation of its gradient which greatly accelerates the convergence of stochastic gradient descent compared to the sigmoid (Tanh) functions.

ReLu output is not zero-centered and could slow down the neural network's performance a bit. But as the authors commented in the article, DN was painful back in 2014 due to the vanishing gradient problem. ReLu is saturated in one direction to assist in anticipating a more resilient result compared to the Tanh function.

Contrasting Practices

How would you quantify if your result duplicated the paper's (hint: what evaluation metric did the researchers use)?

Standard practice in 2014 for Deep Learning would have been more experimental; Google Brain wouldn't release TensorFlow for public use until 2015. The method in which TensorFlow wraps in functions would have had to be designed/programmed on an individual basis for separate applications.

Use of activation functions for hidden layers of a neural net were also different. The paper uses a Tanh activation function. As discussed in the previous section, ReLU would be more appropriate in 2020. Modern approaches use ReLU functions which are generally agreed to have more desirable qualities compared to a sigmoid function. To study the more intricate science behind these functions, Vandit Jain wrote a great article that is found here.

Another contrasting element that would need to be considered is that of the Layer setup. The 2014 approach used a 5 layer 300 unit setup with Tahn activation on all layers; lacking dropout. Dropout helps prevent overfitting. Adding this may have yielded better classification results on the test set.

Weights, momentum and learning rate are also referenced that could be changed with today's technology. The 2014 approach used Stochastic Gradient Descent as its only optimizer function. Stochastic Gradient Descent samples different parts gradient/slope to try to find a global minimal value for error that represents the best solution of weights for an algorithm.

Other options available through TensorFlow today, such as SGD variants ADAM, ADADELTA, and NADAM have different properties that may contribute to better results than SGD. Other optimizers such as RMSprop uses the exponential average of a gradient value by parameter to try to find a weight solution and could possibly outperform SGD with the appropriate parameters.

The metric used in the 2014 approach was the AUC/ROC curve. In today's Neural Net, the theoretical output would also be evaluated with an ROC curve. Given that the task is to classify exotic particles based on sensor data from the Large Hadron Collider, the balance between specificity and sensitivity is important. The paper achieved results of 0.7 to 0.88 AUC. With a more modern architecture, the goal would to performance at least at those levels, but likely above.

Another metric that would adopted would be the F1 score. The case of new particles are severely small compared to the noise(or background) cases. F1 score, which is the balanced metric consisting of Precision and Recall, is a tool used to improve such an imbalanced machine learning classification problem.