ICAI 2023
12th International Conference on Applied Informatics
Eger, Hungary, 2023

What can we learn from Small Data

Tamas Nyiri[a], **Attila Kiss**[ab]

[a]Department of Information Systems, ELTE Eötvös Loránd University, Budapest 1117, Hungary
{nytuaai,kiss}@inf.elte.hu

[b]Department of Informatics, J. Selye University, Komárno 945 01, Slovakia
kissae@ujs.sk

Abstract

Science and technology in the past decade have been completely revolutionized by Deep Learning. From smarter advertisements to self-driving cars, there has been enormous progress on tasks that in the past have seemed completely intractable. But what drove this revolution?

At first glance it would seem like it was all due to the invention of Neural Networks, a Machine Learning technique loosely modeled after the human brain itself. However, this only tells part of the story. Artificial Neural Networks have been around since the 1940's [2] yet for most of their history they haven't been able to surpass traditional techniques.

In reality, the fire was ignited but has been burning on a very low flame for decades. That is because two ingredients were on low supply. For a successful campfire we need flammable material and a healthy supply of oxygen. In our analogy, training data could be viewed as the wood we put on our fire and computational power is the oxygen supply for the training.

The exponential growth of training data and fast computational capabilities of late have been essential for this deep learning revolution to happen. The latter seems to be in constant supply and steadily improving. The former, not always.

There is a multitude of possible scenarios where Deep Learning systems are plagued by a serious lack of training data or quality issues regarding the existing ones.

Just as we have the famous term 'Big Data', which has been coined to describe the new phenomenon of massive data sets fulfilling the criteria of 4 V's (velocity, veracity, volume, and variety) [3], or alternatively (and for our purposes more

fittingly) as characterized by high dimensionality and large sample size [1], we could define a similar concept that we can call 'Small Data'.

In the paper we will try to define Small Data, by first introducing certain scenarios in deep learning that has been characterized as suffering from a low performance due to a lack of data, and working backwards.

We will then shift our focus to the proposed solutions for these scenarios. We will talk about their theoretical background, their pros and cons as well as what might have motivated their use.

Finally, we will summarize our findings and offer some takeaways about this often ignored area of Deep Learning.

References

[1] J. Fan, F. Han, H. Liu: *Challenges of big data analysis*, National science review 1.2 (2014), pp. 293–314.

[2] W. S. McCulloch, W. Pitts: *A logical calculus of the ideas immanent in nervous activity*, The bulletin of mathematical biophysics 5.4 (1943), pp. 115–133.

[3] G. Vossen: *Big data as the new enabler in business and other intelligence*, Vietnam Journal of Computer Science 1.1 (2014), pp. 3–14.