Large Models Reshape AI Research and Applications

Bernd Ludwig[1]

Dear readers,

Fine tuning pretrained models to downstream tasks for a while has been a up to this point rare applied technique to solve prediction tasks in machine learning.

Slowly, in many steps in various domains of machine learning, such as computer vision and natural language processing, became predominant and replaces almost every effort by AI researchers and engineers to specify features for a prediction problem intellectually, i.e. by specifying a theory explaining how decisions for a problem at hand can be taken. This approach was exciting as it could make researchers assume that solving prediction problems was just a matter of collecting a small data set, feeding it into an adapted pretrained model, starting to fine tune it – and that's all. In my lab, we were so keen to believe that objects contained in the training data of pretrained model for computer vision could classifying similar objects in real-time without any fine tuning because they should have very similar features as the objects in the training data. In an experiment with a head mounted augmented reality device we planned to classify the objects in the user's field of view in real-time in order to locate the user in an environment in which other data for tracking the position of persons was hard to obtain.

Unfortunately, the approach did not work well as our doors looked quite different from those in training data. Consequently, fine tuning could not be avoided and took us many hours to record and annotate many video frames. We can now recognize doors in our indoor environment better, but in others, the error rate may not decrease significantly as the doors in our training data as glass doors and seem the vision model to compute other feature values than for more typical doors.

It would help a lot if the model explained those features — currently a hot topic in machine learning with lots of open issues. Britta Wrede in her editorial to a recent issue of this journal about a year ago[1] noted that these phenomena that I described anecdotally above in many AI researchers forms the idea of integrating humans with application knowledge, i.e. on how doors look like, in the process of building models for prediction tasks, among others.

One year sees many days and nights come and disappear again, and many AI publications and demos can be presented in this time. The last year was dominated by OpenAI's chat GPT. At OpenAI, the strategy is "more data is better data", and mixed with some human input to align GPT's predictions to human intentions is sufficient to make GPT produce output that looks reasonable for any input fed into it. Similarly, large vision models[2] can recognize images better than models published previously (we will apply them on our door data to see whether the results are as astonishing as output by chat GPT).

Obviously, these large models change the way in which AI experts do research. Instead of fine tuning, researchers leverage large models with few shot or even zero shot learning in many research tasks.[3] Large models also seem to have learned information about everyday issues humans are faced with. My students (and not only they) use chat GPT to generate program code and ask them questions about the slides I present during my lectures. Typing "chat gpt for" in Google's textbox for their search engine produces auto completions such as "for diet", "for meal planning". I am wondering whether OpenAI at some point of time will publish a sample of the tasks that users confront chat GPT with every day. Also professionals are using chat GPT in medicine[4] or to build interactive assistants.[5] The methodology applied in these research works is completely different from the earlier way and may be used to generate synthetic data in the typical cases of applications with sparse data available. This is

✉ Bernd Ludwig
bernd.ludwig@ur.de

1   Bernd Ludwig Universität Regensburg, Regensburg, Germany

1   https://link.springer.com/article/10.1007/s13218-022-00773-7

2   see e.g. https://arxiv.org/pdf/2212.09408.pdf

3   as described e.g. in https://arxiv.org/abs/2205.11916

4   e.g.   https://www.frontiersin.org/articles/10.3389/frai.2023.1169595/full

5   e.g. https://arxiv.org/abs/2306.16275

kind of a revolution in how research results are obtained and applications are built.

Looking at the rumors about OpenAI's Q* algorithm in these days of November 2023,[6] we have to ask ourselves where this development will and should lead to. Technically, it seems logical that not only language or vision can be processed by identifying patterns and their probability distributions in a given (huge) set of data, but any kind of observations of human behaviour, reasoning, thinking, and inventing as long as it sticks to certain rules or regularities. As a consequence, given the computational power available nowadays, somebody will build models for these observations in order to predict future from past ones. In fact, this is what AI research is doing since it began to exist: trying to formalize the processes underlying these observations in order to be able to act as humans.

We will see debates about how we can keep control of these developments in order to end up as Goethe's Zauberlehrling: "Spirits that I've cited my commands ignore.".[7]

Keep this warning in mind while you enjoy reading this issue of KI,

Bernd Ludwig

---

[6] see https://www.reuters.com/technology/sam-altmans-ouster-openai-was-precipitated-by-letter-board-about-ai-breakthrough-2023-11-22/

[7] Translation taken from https://www.babelmatrix.org/works/de/ (30.11.2023)