Mathematics for and by Large Language Models

Rapport sur les contributions

ID de Contribution: **1**   Type: **Non spécifié**

How can Machine Learning Help Mathematicians

jeudi 23 mai 2024 09:30 (1 heure)

Large Language models have known large successes in recent years. This naturally raises the question: can AI assist mathematicians in solving open problems in mathematics? We will explore how a language model can be trained to learn a mathematical intuition on open problems and guess candidate solutions, with a focus on a few examples. We will also explore the application of LLM to automated theorem proving with an online training procedure and discuss new perspectives in the area.

Orateur: HAYAT, Amaury (École des Ponts ParisTech & CERMICS)

ID de Contribution: **2**   Type: **Non spécifié**

Synthetic Data –Friend or Foe in the Age of Scaling?

jeudi 23 mai 2024 10:30 (1 heure)

As AI and LLM model size grows, neural **scaling laws** have become a crucial tool to predict the improvements of large models when increasing capacity and the size of original (human or natural) training data. Yet, the widespread use of popular models means that the ecosystem of online data and text will co-evolve to progressively contain increased amounts of synthesized data.
In this talk we ask: **How will the scaling laws change in the inevitable regime where synthetic data makes its way into the training corpus?** Will future models, still improve, or be doomed to degenerate up to total **(model) collapse**? We develop a theoretical framework of model collapse through the lens of scaling laws. We discover a wide range of decay phenomena, analyzing loss of scaling, shifted scaling with number of generations, the ''un-learning'' of skills, and grokking when mixing human and synthesized data. Our theory is validated by large-scale experiments with a transformer on an arithmetic task and text generation using the LLM Llama2.

Orateur: KEMPE, Julia (NYU Center for Data Science and Courant Institute of Mathematical Sciences)

ID de Contribution: **3**   Type: **Non spécifié**

A First Approximation to the Mathematical Structure Computed by Large Language Models

jeudi 23 mai 2024 12:00 (1 heure)

Large Language Models are transformer neural networks which are trained to produce a probability distribution on the possible next words to given texts in a corpus, in such a way that the most likely word predicted, is the actual word in the training text.

We will explain what is the mathematical structure defined by such conditional probability distributions of text extensions. Changing the viewpoint from probabilities to -log probabilities we observe that the data of text extensions are encoded in a directed (non-symmetric) metric structure defined on the space of texts $\mathcal{L}$. We then construct a directed metric polyhedron $P(\mathcal{L})$, in which $\mathcal{L}$ is isometrically embedded as generators of certain special extremal rays. Each such generator encodes extensions of a text along with the corresponding probabilities.

Moreover $P(\mathcal{L})$ is $(\min, +)$ (i.e. tropically) generated by the text extremal rays and is the image of a $(\min, +)$ projector (given by the metric on $\mathcal{L}$). This leads to a duality theorem relating the polyhedron $P(\mathcal{L})$ defined by text extensions to one defined by text restrictions. We also explain that the generator of the extremal ray corresponding to a text is approximated by a Boltzmann weighted linear combination of generators of extremal rays corresponding to the words making up that text.

The metric space $\mathcal{L}$ can equivalently be considered as an enriched category and then the embedding into $P(\mathcal{L})$ is the Yoneda embedding into the category of presheaves. In fact all constructions have categorical meaning (in particular generalizing the familiar view of language as a monoid or as a poset with the subtext order). The categorical interpretations will be explained in parallel.
This is joint work with Stéphane Gaubert.

Orateur: VLASSOPOULOS, Yiannis (Athena Research Center & IHES)

ID de Contribution: **4**   Type: **Non spécifié**

Mathematics as a Translation Task - the Importance of Training Distributions

jeudi 23 mai 2024 14:30 (1 heure)

Many problems of mathematics can be set as translation tasks: problems, represented as sentences in some language, are translated into their solutions, by language models trained from synthetic examples. In this setting, we can choose the distribution of problems and solutions we use to train the model. I present examples from three different experiments, which suggest that this can make a large difference in model performance, and provide intuition on the inner workings of transformer models.

Orateur: CHARTON, François (Meta AI Research)

ID de Contribution: **5**   Type: **Non spécifié**

Three Problems in the Mathematics of Deep Learning

jeudi 23 mai 2024 15:30 (1 heure)

Neural networks, particularly LLMs, are notoriously poor at algorithmic tasks, such as sorting, shortest path, and even basic arithmetic. Across three papers, we explored the problem of "aligning" architectures to classical computer programs, and showed that this question relates to familiar mathematical concepts: polynomial functors, cohomology, and higher categories.

Orateur: DUDZIK, Andrew (Google DeepMind)

ID de Contribution: **6** Type: **Non spécifié**

TBA

jeudi 23 mai 2024 17:00 (1 heure)

Orateur: SYNNAEVE, Gabriel (Meta AI Research)