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)