NEURAL INFORMATION PROCESSING SYSTEMS

Algebraic Positional Encodings

Konstantinos Kogkalidis[1,2] Jean-Philipe Bernardy[3,4] Vikas Garg[1,5]

[1] Aalto University
[2] University of Bologna
[3] University of Gotheburg
[4] Chalmers University of Technology
[5] YaiYai Ltd.

TLDR

Algebraic Positional Encodings
Nice & elegant positional encodings for a bunch of different structures.

In a nutshell:

1. A **sensible meta-theory** for positional encodings
2. An **actionable alternative** to the current SOTA

Premise

Attention:

$$\operatorname{atn}(\boldsymbol{X}, \boldsymbol{Y}) := \operatorname{softmax}_{(n)}\left(\frac{(\boldsymbol{X}\boldsymbol{\Phi}^{(q)})(\boldsymbol{Y}\boldsymbol{\Phi}^{(k)})^{\top}}{\sqrt{d}}\right)\boldsymbol{Y}\boldsymbol{\Phi}^{(v)} \tag{1}$$

$\operatorname{atn}(\boldsymbol{X}, \boldsymbol{Y}) \in \mathbb{R}^{m\times d}$ for $\boldsymbol{X} \in \mathbb{R}^{m\times d}$, $\boldsymbol{Y} \in \mathbb{R}^{n\times d}$ and $\Phi^{(q,k,v)} \in \mathbb{R}^{d\times d}$

The problem:

$$\operatorname{atn}(\boldsymbol{X}, \boldsymbol{Y}) \equiv \operatorname{atn}(\boldsymbol{X}, p_n(\boldsymbol{Y})) \tag{2}$$

where $p_n \in \mathcal{S}_n$

Premise

Attention:

$$\operatorname{atn}(\boldsymbol{X}, \boldsymbol{Y}) := \operatorname{softmax}_{(n)}\left(\frac{(\boldsymbol{X}\boldsymbol{\Phi}^{(q)})(\boldsymbol{Y}\boldsymbol{\Phi}^{(k)})^\top}{\sqrt{d}}\right)\boldsymbol{Y}\boldsymbol{\Phi}^{(v)} \tag{1}$$

$\operatorname{atn}(\boldsymbol{X}, \boldsymbol{Y}) \in \mathbb{R}^{m\times d}$ for $\boldsymbol{X} \in \mathbb{R}^{m\times d}$, $\boldsymbol{Y} \in \mathbb{R}^{n\times d}$ and $\Phi^{(q,k,v)} \in \mathbb{R}^{d\times d}$

The problem:

$$\operatorname{atn}(\boldsymbol{X}, \boldsymbol{Y}) \equiv \operatorname{atn}(\boldsymbol{X}, p_n(\boldsymbol{Y})) \tag{2}$$

where $p_n \in \mathcal{S}_n$

Group Algebra 101

What is ... a **group**? 🤔

a **set** S and an **operation** $(\_ \cdot \_) : S \times S \to S$, such that:

- $(a \cdot b) \cdot c = a \cdot (b \cdot c)$
- $a \cdot \epsilon = \epsilon \cdot a = a$
- $a \cdot a^{-1} = a^{-1} \cdot a = \epsilon$

Trivia:

- If the whole group can be expressed using only $S_0 \subset S$, then $G = \langle S_0 \rangle$; G is **generated** by S_0.

Group Algebra 101

What is ... a **group**? 🤔

a **set** S and an **operation** $(\_ \cdot \_) : S \times S \to S$, such that:

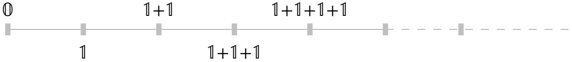
- $(a \cdot b) \cdot c = a \cdot (b \cdot c)$
- $a \cdot \epsilon = \epsilon \cdot a = a$
- $a \cdot a^{-1} = a^{-1} \cdot a = \epsilon$

Trivia:

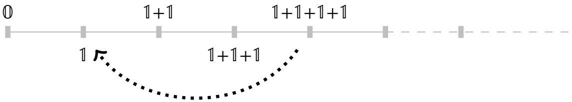
- If the whole group can be expressed using only $S_0 \subset S$, then $G = \langle S_0 \rangle$; G is **generated** by S_0.

Points on a sequence $\equiv \mathbb{N}$

Points on a sequence $\equiv \mathbb{N}$



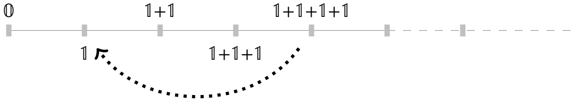
Paths on a sequence $\equiv \mathbb{Z}$



$$-(\mathbb{1}+\mathbb{1}+\mathbb{1}+\mathbb{1})+\mathbb{1}$$
$$=(-\mathbb{1})+(-\mathbb{1})+(-\mathbb{1})+(-\mathbb{1})+\mathbb{1}$$
$$=(-\mathbb{1})+(-\mathbb{1})+(-\mathbb{1})$$

trivia #1: $\mathbb{Z} := \langle \mathbb{1} \rangle$

Paths on a sequence $\equiv \mathbb{Z}$



$$
\begin{aligned}
&-(\mathbb{1}+\mathbb{1}+\mathbb{1}+\mathbb{1})+\mathbb{1}\\
&=(-\mathbb{1})+(-\mathbb{1})+(-\mathbb{1})+(-\mathbb{1})+\mathbb{1}\\
&=(-\mathbb{1})+(-\mathbb{1})+(-\mathbb{1})
\end{aligned}
$$

trivia #1: $\mathbb{Z} := \langle \mathbb{1} \rangle$

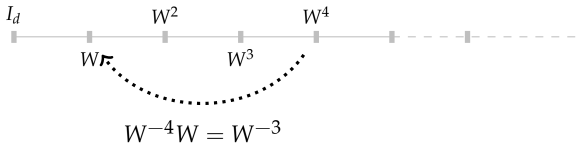
Paths on a sequence $\equiv \mathbb{Z}$

Consider $\phi : \mathbb{Z} \to O(d)$, such that:

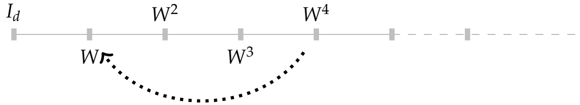
$$\begin{aligned} \phi(1) &\mapsto W \in O(d) \\ \phi(m+n) &\mapsto \phi(m)\phi(n) \\ \phi(-m) &\mapsto \phi(m)^\top \equiv \phi(m)^{-1} \end{aligned}$$

trivia #2: orthogonal matrices form a group under multiplication

Paths on a sequence $\equiv \mathbb{Z}$



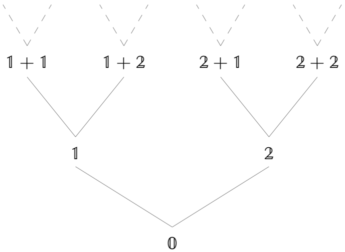
Paths on a sequence $\equiv \mathbb{Z}$



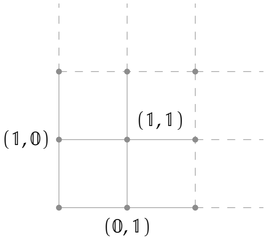
$$W^{-4}W = W^{-3}$$

.. looks familiar?

Paths on many other structures are also groups



$$\phi(\mathbb{1}) \mapsto L$$
$$\phi(\mathbb{2}) \mapsto R$$



$$\phi((m, n)) \mapsto \phi(m) \oplus \phi(n)$$

📙 **Preprint**
arxiv.org/abs/2312.16045

🍁 **Poster Session**
Wed 11 Dec, 11 a.m - 2 p.m. PST @ NeurIPS 2024

⭐ **Code**
github.com/konstantinosKokos/ape

🗣️ Get in touch!
Questions/comments/feedback welcome.

Thank you!