Numbers Already Carry Their Own Embeddings

Suhyun Bae and Donghun Lee
Department of Mathematics, Korea University

NEURAL INFORMATION PROCESSING SYSTEMS

AIML@K

Motivation

Solve: 5.9 = x + 5.11

Let's solve step by step:

$$5.9 = x + 5.11$$

Subtract 5.11 from both sides:

$$5.9 - 5.11 = x$$

Compute carefully:

$$5.90 - 5.11 = 0.79 - 0.00 = -0.21$$

Answer:

$$x = -0.21$$

Even LLMs often fail at basic arithmetic

Approach

Problem

- Numbers are tokenized and embedded arbitrarily during training

"123" ➡ "1" "2" "3"

- Embeddings don't inherently capture number relationships

Embed(1) ➕ Embed(2) ≠ Embed(3)

Hypothesis

- **Preserving algebraic structure of numbers in embedding space** is the solution

Adele Ring

Definition

- Let $\mathbb{Q}_p$ p-adic numbers and $\mathbb{Z}_p$ p-adic integers for some prime p Then Adele ring $\mathbb{A}_\mathbb{Q}$ is defined as:

$$\mathbb{A}_\mathbb{Q} \stackrel{\text{def}}{=} \mathbb{R} \times \prod_p (\mathbb{Q}_p, \mathbb{Z}_p)$$

- Let $A: \mathbb{Q} \to \mathbb{A}_\mathbb{Q}$. Then, for any rational number q,

$$A(q) = (q,\ q_2,\ q_3,\ \dots,\ q_p,\ \dots)$$

- For example, $\frac{7}{5}$ can be represented as:

$$\frac{7}{5} \to (1.4,\ \dots 11011_2,\ \dots 01212_3,\ 1.2_5,\ \dots 54130_7,\ \dots)$$

Key Features

- **Maps any rational number to Adele ring directly from number theory**
- Element-wise **Additive**

$$A(q_1 + q_2) = A(q_1) \oplus A(q_2)$$

- Element-wise **Multiplicative**

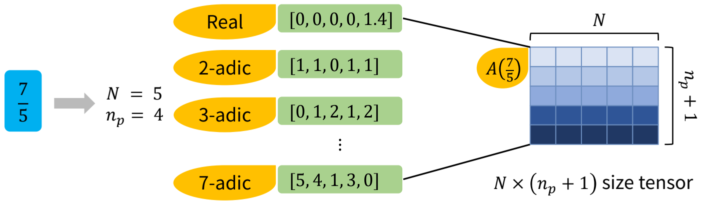
$$A(q_1 \times q_2) = A(q_1) \otimes A(q_2)$$

Where $\oplus$, $\otimes$ denotes element-wise operations

AOE : Our Method

AOE (Adelic Operation-preserved Embeddings)

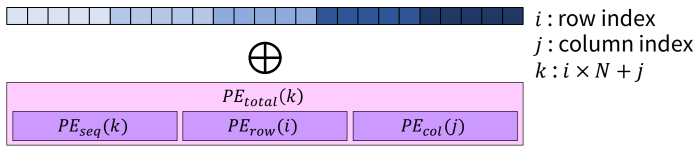
- Truncate Adele representation to **practical N-digit precision**
- Each number is **encoded as a tensor of shape $N \times (n_p + 1)$** where n_p means the number of primes



- **Flat tensor and apply 2D positional encoding**

PE is standard sinusoidal encoding

$$PE_{total}(k) = PE_{seq}(k) \oplus PE_{row}(i) \oplus PE_{col}(j)$$



Experiment & Result

Model

- 6-layer Transformer Encoder
- **Baseline:** Trainable "`nn.Embedding`"+ 1D PE
- **Ours (AOE):** Training-free AOE + 2D PE

Benchmark

- **Algebraic Combinatorics Dataset (ACD)**
- **Dataset of research-level problems in algebraic combinatorics**
- Task Example (mHeight Function)

➢ $x = [2, 1, 5, 0, 7, 3, 6, 4],\ \ y = 0$

Result

Dataset		Accuracy ↑	
Task	**n**	**Baseline**	**AOE(Ours)**
Lattice paths	10	66.19	70.10
	11	66.30	66.30
	12	66.50	78.44
Weaving patterns	6	53.02	100.00
	7	51.53	100.00
Quiver mutation classes		45.13	91.11
Grassmannian cluster algebras		91.27	97.39
Schubert polynomials	4	50.59	87.89
	5	49.83	96.87
	6	50.00	99.96
mHeight	8	91.42	96.94
	9	93.20	99.17
	10	94.15	99.66

AOE outperforms the baseline across all tasks

Discussion

1. Restricted to rational numbers ($\mathbb{Q}$) by Definition
   ▶ Extend AOE to irrationals, complex values
2. Does not handle numbers written as text (e.g., "Three hundred and fifty")
   ▶ Align AOE representation with text embedding spaces

* This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2025-24873052)