

PE Diagram on changes with position

۴- تحلیل رفتار نمودارها

۴.۱. برای $i = 0$

- کوچکترین denom
- بیشترین فرکانس
- نوسانات بسیار سریع نسبت به position

تعییر:

این بعد اطلاعات محلی و کوتاهبرد موقعیت را کد می‌کند.

۴.۲. برای $i = 2$

- کمی فرکانس کمتر
- همچنان نوسانات واضح

۴.۳. برای $i = 8$

- فرکانس به طور محسوسی کاهش می‌یابد
- موج نرمتر و کشیده‌تر

۴.۴. برای $i = 32$

- بسیار کم فرکانس
- تغییرات آهسته در طول position

تعییر:

این بعد اطلاعات بلندبرد و کلی از موقعیت را نمایش می‌دهد.

۵. تفسیر مفهومی (نکته بسیار مهم)

ترکیب این ابعاد مختلف باعث می‌شود:

- هر موقعیت (pos) یک بردار یکتا داشته باشد

مدل بتواند:

- فاصله نسبی توکن‌ها را یاد بگیرد
- الگوهای محلی و سراسری ترتیب را همزمان در کد بتواند از اختلاف فاز سینوس‌ها فاصله را استنتاج کند
- Attention

بهصورت شهودی:

- ابعاد کوچک $i \Rightarrow$ جزئیات دقیق ترتیب
- ابعاد بزرگ $i \Rightarrow$ ساختار کلی توالی

جمع‌بندی

- این کد رفتار **Positional Encoding** سینوسی را برای ابعاد مختلف embedding نمایش می‌دهد.
- با افزایش i ، فرکانس موج کاهش می‌یابد.
- این طراحی چندمقیاسی، دلیل توانایی Transformer در در ک ترتیب و فاصله توکن‌هاست.
- نمودارها نشان می‌دهند که PE ترکیبی از اطلاعات محلی و سراسری را به مدل تزریق می‌کند.

PE on position های ثابت

۱. رفتار کلی در طول ابعاد

برای هر $:pos$

- در ابعاد پایین (i کوچک):
 - نوسانات سریع و شدید
- در ابعاد بالا (i بزرگ):
 - نوسانات آهسته و صاف

این یعنی:

بهصورت طیف فرکانسی از اطلاعات مکانی ساخته شده است.

۲. مقایسه موقعیت‌های مختلف

$$pos = 0.5 \cdot i$$

- $angles = 0$
- $\sin(\theta) = 0, \cos(\theta) = 1$
- الگوی کاملاً ساختاریافته و ثابت

بردار PE در این حالت:

- مستقل از فرکانس
- صرفاً نابع ابعاد parity

pos = 1.5.2.2

- اختلاف فاز جزئی نسبت به $pos=0$
- تغییرات بیشتر در ابعاد با فرکانس بالا

pos = 10.5.2.3

- جابه‌جایی فاز محسوس
- تغییرات در اکثر ابعاد قابل مشاهده

100 و pos = 50.5.2.4

- تغییر فاز بزرگ
- در ابعاد پایین ممکن است چندین چرخه کامل طی شود
- در ابعاد بالا هنوز تغییرات نرم هستند

۶. برداشت مفهومی بسیار مهم

این نمودارها نشان می‌دهند که:

1. هر **position** یک امضا طیفی یکتا دارد
 2. اختلاف دو **position** برابر است با:
 - اختلاف فاز در تمام فرکانس‌ها
 3. می‌تواند از طریق:
 - Self-Attention
 - dot-product
 - sin/cos ترکیب
- فاصله نسبی توکن‌ها را استنتاج کند

به صورت شهودی:

- ابعاد پایین → حساس به تغییرات کوچک **position**
- ابعاد بالا → مقاوم به تغییرات کوچک و مناسب برای فاصله‌های بزرگ

۷. ارتباط مستقیم با Attention

اگر دو توکن با position‌های متفاوت را در نظر بگیریم:

- شباهت بردارهای PE تابعی از اختلاف position است
- این باعث می‌شود Attention بتواند:

- ترتیب
 - فاصله
 - جهت (قبل/بعد)
- را بدون یادگیری صریح بفهمد

جمع‌بندی

- این کد ساختار داخلی بردار Positional Encoding را برای position‌های ثابت نمایش می‌دهد.
- هر بردار PE ترکیبی از \sin/\cos با فرکانس‌های مختلف است.
- تغییر position معادل تغییر فاز در کل طیف فرکانسی embedding است.
- این طراحی امکان استنتاج فاصله و ترتیب را به Transformer می‌دهد بدون نیاز به RNN یا CNN.