

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

تعبیر:

این بُعد 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 به‌طور محسوسی کاهش می‌یابد
- موج نرم‌تر و کشیده‌تر

۴.۴. برای $i = 32$

- بسیار کم‌فرکانس
- تغییرات آهسته در طول position

تعبیر:

این بُعد 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 درک کند
- Attention بتواند از اختلاف فاز سینوس‌ها فاصله را استنتاج کند

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 درک ترتیب و فاصله توکن‌هاست.
- نمودارها نشان می‌دهند که PE ترکیبی از اطلاعات محلی و سراسری را به مدل تزریق می‌کند.

های ثابت PE on position

۵.۱. 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:

embedding به‌صورت طیف فرکانسی از اطلاعات مکانی ساخته شده است.

۵.۲. مقایسه موقعیت‌های مختلف

$$pos = 0 \quad ۵.۲.۱$$

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

بردار PE در این حالت:

- مستقل از فرکانس
- صرفاً تابع parity ابعاد

۵.۲.۲. $pos = 1$

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

۵.۲.۳. $pos = 10$

- جابجایی فاز محسوس
- تغییرات در اکثر ابعاد قابل مشاهده

۵.۲.۴. $pos = 50$ و 100

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

به صورت شهودی:

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

۷. ارتباط مستقیم با Attention

اگر دو توکن با position های متفاوت را در نظر بگیریم:

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

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

جمع بندی

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