

# مقایسه مدل‌های Seq2Seq مبتنی بر LSTM و Transformer

نام گروه

Tensor Titans

اعضای گروه

غزل عسکری

یاسمین صرافی

سپیده سلیمانیان

امیرحسین رجبی

محمدرضا ویلانی

تاریخ تنظیم سند

شهریور ۱۴۰۳

در این پروژه، یک مدل Seq2Seq مبتنی بر Transformer برای ترجمه انگلیسی به فارسی پیاده‌سازی شده است. مدل‌های Seq2Seq به‌طور سنتی با استفاده از معماری LSTM (نوعی شبکه عصبی بازگشتی یا RNN) ساخته می‌شدند.

## مدل Seq2Seq مبتنی بر LSTM

حافظه طولانی کوتاه مدت (LSTM - Long Short Term Memory) نوعی شبکه عصبی بازگشتی است که به دلیل قابلیت غلبه بر مشکل ناپدید شدن گرادیان، برای مدیریت دنباله‌های طولانی بسیار موثر است. مدل‌های Seq2Seq مبتنی بر LSTM به صورت زیر عمل می‌کنند:

1. رمزگذار (Encoder): یک LSTM دنباله ورودی (مانند یک جمله انگلیسی) را به‌صورت توکن به توکن پردازش می‌کند و یک بردار حالت پنهان (context vector) ایجاد می‌کند که خلاصه‌ای از ورودی است.
2. رمزگشا (Decoder): رمزگشا که آن هم یک LSTM است، از این بردار حالت پنهان برای تولید دنباله خروجی (مانند ترجمه فارسی) توکن به توکن استفاده می‌کند.

### ویژگی‌های کلیدی مدل‌های مبتنی بر LSTM:

- پردازش ترتیبی: این مدل‌ها توکن‌ها را به صورت دنباله‌وار و یک به یک پردازش می‌کنند.
- بردار کانتکست (Context Vector): بردار کانتکست با اندازه ثابت ممکن است اطلاعات حیاتی را برای دنباله‌های طولانی از دست بدهد.
- وابستگی‌های طولانی مدت: LSTM‌ها برای نگهداری و به‌روزرسانی حافظه در دنباله‌های طولانی طراحی شده‌اند، اما با وجود این همچنان در مدیریت وابستگی‌های بسیار طولانی مشکل دارند.
- مکانیزم توجه (Attention): مکانیزم توجه می‌تواند به مدل اضافه شود تا رمزگشا در هنگام تولید خروجی، روی بخش‌های مرتبط از دنباله ورودی تمرکز کند.

## مدل Seq2Seq مبتنی بر Transformer

در مقابل، مدل‌های Transformer نیازی به بازگشت (recurrence) ندارند و از یک مکانیزم توجه (self-attention) برای پردازش دنباله‌ها به صورت موازی و بهبود درک وابستگی‌ها بین توکن‌ها استفاده می‌کنند. معماری این مدل شامل:

1. رمزگذار (Encoder): از چندین لایه توجه خودکار (self-attention) برای پردازش همزمان تمامی توکن‌ها استفاده می‌کند. این مکانیزم باعث می‌شود تا مدل بتواند روابط بین کلمات را در دنباله ورودی بهتر درک کند و به بخش‌های مختلف جمله به‌طور همزمان توجه کند.
2. رمزگشا (Decoder): مشابه رمزگذار است اما علاوه بر دنباله خروجی، به خروجی رمزگذار نیز توجه می‌کند تا ترجمه را تولید کند.

## مقایسه

ویژگی	Seq2Seq مبتنی بر LSTM	Seq2Seq مبتنی بر Transformer
نوع معماری	شبکه عصبی بازگشتی (RNN)	Transformer (مبتنی بر توجه خودکار)
پردازش ترتیبی	بله، توکن به توکن	خیر، پردازش موازی
مدیریت کانتکست	استفاده از بردار کانتکست ثابت	استفاده از توجه خودکار برای کانتکست پویا
مکانیزم توجه (Attention)	به‌صورت جداگانه اضافه می‌شود	مکانیزم توجه چندگانه داخلی
مدیریت دنباله‌های طولانی	خوب، اما مشکل در وابستگی‌های خیلی طولانی	بسیار موثر، مدیریت طبیعی وابستگی‌های طولانی
موازی‌سازی	محدود (پردازش توکن به توکن)	کاملاً موازی، پردازش سریع‌تر
اطلاعات موقعیت	نیازی ندارد (ترتیب در معماری ذاتی است)	نیاز به رمزگذاری موقعیت (Positional Encoding)
زمان آموزش	آهسته‌تر به دلیل پردازش ترتیبی	سریع‌تر به دلیل پردازش موازی
قابلیت مقیاس‌پذیری	کمتر مقیاس‌پذیر	بسیار مقیاس‌پذیر
پیچیدگی مدل	ساده‌تر اما کم‌کارآمدتر برای دنباله‌های طولانی	پیچیده‌تر اما موثرتر برای دنباله‌های طولانی
استفاده از حافظه	کمتر استفاده‌کننده از حافظه	نیاز بیشتر به حافظه به دلیل لایه‌های توجه

کیفیت ترجمه	خوب، اما ممکن است در دنباله‌های طولانی یا پیچیده کاهش یابد	کیفیت بالا، به ویژه برای دنباله‌های طولانی و پیچیده
-------------	--	---

## نکات کلیدی مقایسه

### 1. مدیریت دنباله‌های طولانی:

- LSTMها در به خاطر سپردن اطلاعات در دنباله‌های طولانی خوب عمل می‌کنند، اما در هنگام برخورد با وابستگی‌های بسیار دور یا پیچیده، ممکن است عملکردشان افت کند. اضافه کردن مکانیزم توجه می‌تواند این مشکل را تا حدودی حل کند.
- در مقابل، Transformerها با استفاده از توجه خودکار، می‌توانند وابستگی‌ها را در هر فاصله‌ای در دنباله به‌طور موثری مدیریت کنند و برای دنباله‌های طولانی مناسب‌تر هستند.
- در نتیجه، Transformerها به دلیل توانایی در پردازش موازی و استفاده از مکانیزم توجه خودکار، حتی در توالی‌های کوتاه نیز عملکرد بهتری نسبت به LSTMها دارند و قادر به تشخیص وابستگی‌های محلی و جهانی در یک دنباله هستند.
- با پیشرفت‌های جدید در زمینه مدل‌های پیش‌آموزش مانند BERT و GPT، که هر دو مبتنی بر معماری Transformer هستند، عملکرد مدل‌های ترجمه و تولید متن به میزان قابل توجهی بهبود یافته است. LSTMها در این زمینه محدودتر بوده و در حجم داده‌های عظیم کمتر به کار می‌روند.

### 2. مکانیزم اتنشن (Attention):

- در مدل‌های مبتنی بر LSTM، توجه به‌صورت جداگانه اضافه می‌شود و معمولاً در هنگام رمزگشایی بر روی یک مجموعه ثابت از توکن‌های ورودی تمرکز می‌کند.
- در مدل‌های Transformer، مکانیزم توجه چندگانه (multi-head attention) به‌طور ذاتی در معماری وجود دارد و این امکان را فراهم می‌کند که مدل به‌طور همزمان به چندین بخش از دنباله ورودی توجه کند، که باعث بهبود کیفیت ترجمه می‌شود.

### 3. موازی‌سازی:

- LSTMها دنباله‌ها را به‌صورت ترتیبی پردازش می‌کنند که باعث می‌شود آموزش و استنتاج (تولید) کندتر شود، زیرا هر توکن باید به ترتیب پردازش شود.
- در مقابل، Transformerها امکان پردازش موازی تمام توکن‌ها را فراهم می‌کنند که این امر سرعت آموزش را به‌طور قابل توجهی افزایش می‌دهد و مقیاس‌پذیری مدل را بهتر می‌کند.

### 4. اطلاعات موقعیت:

- در LSTMها ترتیب توکن‌ها به‌طور ذاتی مدیریت می‌شود و نیازی به اطلاعات اضافی درباره ترتیب توکن‌ها وجود ندارد.

- در Transformer ها، چون توکن‌ها به‌طور موازی پردازش می‌شوند، رمزگذاری موقعیت (Positional Encoding) ضروری است تا مدل بتواند ترتیب توکن‌ها را درک کند.
- 5. کارایی آموزش:
- LSTM ها به دلیل ماهیت ترتیبی‌شان به زمان بیشتری برای آموزش نیاز دارند. حتی با بهینه‌سازی‌های مدرن، آن‌ها کندتر از Transformer ها هستند.
- Transformer ها به دلیل پردازش موازی سریع‌تر آموزش می‌بینند، اما به دلیل استفاده از لایه‌های توجه به حافظه بیشتری نیاز دارند.
- 6. کیفیت ترجمه:
- LSTM ها می‌توانند ترجمه‌های خوبی ارائه دهند، اما کیفیت آن‌ها ممکن است در جمله‌های طولانی یا پیچیده افت کند، مگر اینکه مکانیزم توجه اضافه شود.
- Transformer ها به طور کلی کیفیت ترجمه بالاتری ارائه می‌دهند، به‌ویژه برای جمله‌های طولانی یا پیچیده، به دلیل قابلیت تمرکز روی تمام بخش‌های مرتبط دنباله با استفاده از توجه چندقابله.
- علاوه بر ترجمه، معماری Transformer در مدل‌های مولدی مانند GPT نیز بسیار موفق بوده است. این مدل‌ها می‌توانند متن‌های طولانی و پیچیده تولید کنند و کیفیت بالایی در تولید محتوا از خود نشان دهند. در مقابل، LSTM ها در تولید متن‌های طولانی ممکن است با مشکلاتی همچون از دست دادن وابستگی‌های دورتر مواجه شوند.

## نتیجه‌گیری

در حالی که مدل‌های Seq2Seq مبتنی بر LSTM پایه و اساس وظایفی مانند ترجمه را گذاشتند، مدل‌های مبتنی بر Transformer پیشرفت‌های قابل توجهی را به ارمغان آوردند. مکانیزم توجه خودکار در Transformer ها به آن‌ها امکان می‌دهد تا دنباله‌های طولانی را به‌طور موثرتری مدیریت کنند، توکن‌ها را به‌طور موازی پردازش کنند و ترجمه‌هایی با کیفیت بالاتر تولید کنند. این ویژگی‌ها مدل‌های Transformer را برای وظایف مقیاس بزرگ مانند ترجمه ماشین ایده‌آل می‌سازد. با این حال، LSTM ها هنوز برای داده‌های کوچکتر یا وظایف ساده‌تر گزینه مناسبی هستند.