خلاصهٔ دوره‌های هوش مصنوعی گوگل — یادداشت مطالعه به فارسی

معماری مدل‌های مدرن: توجه، رمزگذار-رمزگشا، ترنسفورمر و BERT

Attention Mechanism · Encoder-Decoder · Transformer & BERT · Image Captioning

ارائه‌دهنده: Google Cloud / Google Skills

ادغام ۴ دورهٔ کوتاه در یک فصل

سطح: پیشرفته — نیازمند پیش‌زمینهٔ ML

فایل ۱۷ از ۲۲ — تهیه‌شده برای تدریس

موضوع: معماری‌هایی که زیربنای LLMهای امروزی‌اند **تعداد درس:** ۴ دورهٔ کوتاه (ادغام‌شده) **سطح:** پیشرفته

پیش‌نیاز: شبکهٔ عصبی، تعبیه‌ها، پایتون/TensorFlow

چرا این چهار دوره در یک فایل آمده‌اند

این چهار دورهٔ کوتاه در مسیر **Advanced: Generative AI for Developers** پشت سر هم می‌آیند و در واقع یک روایت واحد را می‌سازند: چطور از شبکه‌های عصبی ساده به معماری‌ای رسیدیم که امروز پشت هر مدل زبانی بزرگی است. جدا خواندنشان معنا را تکه‌تکه می‌کند.

بخش یک — سازوکار توجه (Attention Mechanism)

مسئله‌ای که حل می‌کند

- مدل‌های دنباله‌ای قدیمی (RNN و LSTM) کل جمله را در یک بردار ثابت فشرده می‌کردند؛ در جمله‌های بلند، اطلاعات ابتدای جمله گم می‌شد.
- همچنین پردازش **ترتیبی** بود و امکان موازی‌سازی نداشت — یعنی آموزش کند.

ایدهٔ توجه

- به‌جای فشرده کردن همه‌چیز در یک بردار، مدل در هر گام تصمیم می‌گیرد به **کدام بخش‌های ورودی** بیشتر توجه کند.
- هر توجه یک **وزن** دارد؛ مجموع وزن‌ها مشخص می‌کند خروجی این گام بیشتر متأثر از کدام کلمات است.
- نتیجه: مدل می‌تواند رابطهٔ کلماتی را که در جمله از هم دورند، مستقیم ببیند.

> **شهود سادهٔ توجه**
>
> وقتی «او کتاب را برداشت چون **آن** سنگین بود» را می‌خوانید، ذهنتان «آن» را به «کتاب» وصل می‌کند نه به «او». توجه دقیقاً همین وصل کردن است — و مدل یاد می‌گیرد وزنش را خودش تنظیم کند.

بخش دو — معماری رمزگذار-رمزگشا (Encoder-Decoder)

- معماری استاندارد برای وظایف **دنباله به دنباله**: ترجمه، خلاصه‌سازی، پرسش‌وپاسخ.

- **رمزگذار** ورودی را می‌خواند و به یک نمایش برداری معنادار تبدیل می‌کند.
- **رمزگشا** از آن نمایش، خروجی را **توکن به توکن** تولید می‌کند.
- افزودن توجه به این معماری، جهش کیفی بزرگی ایجاد کرد.
- دوره شامل پیاده‌سازی عملی یک مدل ساده در TensorFlow است.

بخش سه — ترنسفورمر و BERT

ترنسفورمر

- معماری‌ای که **کاملاً بر پایهٔ توجه** ساخته شد و بازگشتی (recurrence) را حذف کرد.
- مزیت اصلی: **موازی‌سازی کامل** در آموزش — همین امکان ساخت مدل‌های عظیم امروزی را داد.
- اجزای کلیدی: **خودتوجهی چندسر** (multi-head self-attention)، **رمزگذاری موقعیت** (positional encoding)، لایه‌های پیش‌خور، نرمال‌سازی لایه و اتصال‌های میان‌بر.
- نکته: چون ترتیب ذاتی وجود ندارد، **رمزگذاری موقعیت** به مدل می‌گوید هر کلمه کجای جمله است.

BERT

- سرنام **Bidirectional Encoder Representations from Transformers**.
- فقط از بخش **رمزگذار** ترنسفورمر استفاده می‌کند.
- **دوسویه** است: هر کلمه را با توجه به کلمات قبل **و بعد** از خودش می‌فهمد.
- دو وظیفهٔ پیش‌آموزش:
  - **مدل‌سازی زبان نقاب‌دار** (Masked LM): بخشی از کلمات پنهان می‌شود و مدل باید حدس بزند.
  - **پیش‌بینی جملهٔ بعدی**: آیا این دو جمله پشت سر هم می‌آیند؟
- کاربرد: فهم متن — طبقه‌بندی، تحلیل احساس، پاسخ به پرسش. برای **تولید** متن مناسب نیست (آن کار مدل‌های رمزگشا-محور است).

بخش چهار — ساخت مدل زیرنویس تصویر (Image Captioning)

- تمرین جمع‌بندی: ترکیب **بینایی** و **زبان** در یک مدل.
- ساختار: یک **رمزگذار تصویری** (شبکهٔ عصبی پیچشی) ویژگی‌های تصویر را استخراج می‌کند، یک **رمزگشای زبانی** با کمک **توجه** جمله را تولید می‌کند.

- پیاده‌سازی عملی و آموزش مدل روی Google Cloud.
- ارزش آموزشی: نشان می‌دهد همان مفاهیم توجه و رمزگذار-رمزگشا، فراتر از متن هم کار می‌کنند — و همین ریشهٔ **مدل‌های چندوجهی** امروزی است.

جمع‌بندی سریع برای مرور

- خط روایت: **RNN** ← **افزودن توجه** ← **حذف بازگشت و ماندن با توجه (ترنسفورمر)** ← **تخصصی‌سازی (BERT) برای فهم، مدل‌های رمزگشا برای تولید)** ← **گسترش به چندوجهی**.
- «Attention is all you need» شعار تبلیغاتی نبود؛ توصیف دقیق معماری بود.
- BERT برای **فهم**، مدل‌های رمزگشا-محور برای **تولید** — این تمایز در انتخاب مدل عملی مهم است.
- بدون فهم این فصل، بحث‌هایی مثل پنجرهٔ زمینه، هزینهٔ محاسباتی توجه و محدودیت‌های طول ورودی مبهم می‌مانند.

منابع

- https://skills.google/paths/183
- https://cloud.google.com/blog/topics/training-certifications/new-generative-ai-trainings-from-google-cloud
- https://www.shiksha.com/online-courses/transformer-models-and-bert-model-course-googl118