



سه مرحله اصلی کدنویسی یک مدل زبان بزرگ (LLM) عبارت‌اند از: پیاده‌سازی معماری LLM و فرایند آماده‌سازی داده‌ها (مرحله ۱)، پیش‌آموزش یک LLM برای ایجاد یک مدل پایه (مرحله ۲) و تنظیم دقیق مدل پایه برای تبدیل آن به یک دستیار شخصی یا طبقه‌بند متن (مرحله ۳). هر یک از این مراحل در این کتاب بررسی و پیاده‌سازی می‌شوند.

نویسنده: سباستین راشکا

مدل‌های زبانی بزرگ

از ایده تا پیاده‌سازی

مترجم:

مجتبی کردآبادی

(دانشجوی دکتری مهندسی کامپیوتر - هوش مصنوعی)

مدل‌های زبانی بزرگ از ایده تا پیاده‌سازی

مترجم: مجتبی کردآبادی
ویراستار علمی: دکتر رامین مولاناپور
ناشر: انتشارات آتی‌نگر
مدیر هنری و طراح جلد: همتا بیداریان
چاپ اول، ۱۴۰۴
شمارگان: ۱۰۰ جلد
قیمت: ۵,۹۰۰,۰۰۰ ریال
شابک: ۹۷۸-۶۲۲-۸۲۴۵-۶۱-۴

ISBN: 978-622-8245-61-4

حق چاپ برای انتشارات آتی‌نگر محفوظ است.

نشانی دفتر فروش: خیابان جمالزاده جنوبی، روبه‌روی کوچه رشتچی، پلاک ۱۴۴، واحد ۱

نمبر: ۶۶۵۶۵۳۳۷

تلفن: ۸-۶۶۵۶۵۳۳۶



www.ati-negar.com



info@ati-negar.com



@atinegar1396



publication_atinegar

سرشناسه: راشکا، سباستین، 1986، Raschka، Sebastian - ۱۹۸۶

مدل‌های زبانی بزرگ «از ایده تا پیاده‌سازی» / نویسنده: سباستین راشکا / مترجم: مجتبی کردآبادی / ویراستار علمی: رامین مولاناپور

تهران: آتی‌نگر، ۱۴۰۴

۴۲۶ ص: مصور، جدول، نمودار.

ISBN: 978-622-8245-61-4

فیبا.

یادداشت: عنوان اصلی کتاب: Build a large language model (from scratch), 2025.

موضوع: هوش مصنوعی - Artificial intelligence

موضوع: پردازش زبان طبیعی - Natural language processing (Computer science)

شناسه افزوده: کردآبادی، مجتبی، ۱۳۶۵، - مترجم

شناسه افزوده: مولاناپور، رامین، ۱۳۵۲، - ویراستار علمی

شناسه افزوده: بیداریان، همتا، ۱۳۶۱، - مدیر هنری

رده‌بندی کنگره

رده‌بندی دبویی

شماره کتابشناسی ملی

Q۳۳۵

۰۰۶/۳

۱۰۰۷۱۴۳۹

فهرست مطالب

پیشگفتار.....	۹
درباره این کتاب.....	۱۱
چه کسانی باید این کتاب را بخوانند.....	۱۳
نحوه سازماندهی این کتاب: یک نقشه راه.....	۱۴
درباره کد.....	۱۵
سایر منابع آنلاین.....	۱۶
در مورد نویسنده.....	۱۷
در مورد تصویر روی جلد.....	۱۹
فصل اول: درک مدل‌های زبانی بزرگ.....	۲۱
۱-۱ مدل زبانی بزرگ چیست؟.....	۲۳
۱-۲ کاربردهای مدل‌های زبانی بزرگ.....	۲۵
۱-۳ مراحل ساخت و استفاده از مدل زبانی بزرگ.....	۲۷
۱-۴ معرفی معماری ترنسفورمر.....	۲۹
۱-۵ استفاده از مجموعه داده‌های بزرگ.....	۳۴
۱-۶ نگاهی دقیق‌تر به معماری GPT.....	۳۶
۱-۷ ساخت یک مدل زبانی بزرگ.....	۳۸
خلاصه.....	۳۹
فصل دوم: کار با داده‌های متنی.....	۴۱
۲-۱ درک تعبیه‌های کلمه.....	۴۲
۲-۲ توکن‌سازی متن.....	۴۵

۲-۳	تبدیل توکن‌ها به شناسه‌های توکن	۵۰
۲-۴	افزودن توکن‌های زمینه خاص	۵۵
۲-۵	رمزگذاری زوج بایت	۵۹
۲-۶	نمونه‌برداری داده با یک پنجره کشویی	۶۲
۲-۷	ایجاد تعبیه‌های توکن	۷۰
۲-۸	رمزگذاری موقعیت کلمات	۷۳
۷۸	خلاصه	

فصل سوم: کدنویسی سازوکارهای توجه

۳-۱	مشکل مدل‌سازی دنباله‌های طولانی	۸۳
۳-۲	ثابت وابستگی‌های داده با سازوکارهای توجه	۸۵
۳-۳	توجه به بخش‌های مختلف ورودی با خودتوجه	۸۶
۳-۴	پیاده‌سازی خودتوجهی با وزن‌های آموزش‌پذیر	۹۷
۳-۵	پنهان کردن کلمات آینده با توجه علی	۱۰۹
۳-۶	گسترش توجه تک‌سر به توجه چندسر	۱۱۹
۱۳۰	خلاصه	

فصل چهارم: پیاده‌سازی یک مدل GPT

۱۳۱	برای تولید متن (از پایه)	۱۳۱
۴-۱	کدنویسی معماری مدل زبانی بزرگ	۱۳۲
۴-۲	نرمال‌سازی فعال‌سازها با نرمال‌سازی لایه	۱۴۰
۴-۳	پیاده‌سازی یک شبکه پیشخور با فعال‌سازی‌های GELU	۱۴۷
۴-۴	افزودن اتصالات میانبر	۱۵۲
۴-۵	اتصال لایه‌های توجه و لایه خطی در یک بلوک ترنسفورمر	۱۵۶
۴-۶	کدنویسی مدل GPT	۱۶۰
۴-۷	تولید متن	۱۶۷
۱۷۲	خلاصه	

فصل پنجم: پیش‌آموزش روی داده‌های بدون برچسب

۵-۱	ارزیابی مدل‌های متن مولد	۱۷۶
-----	--------------------------	-----

۱۹۵	۵-۲ آموزش یک مدل زبانی بزرگ.....
۲۰۲	۵-۳ استراتژی‌های رمزگشایی برای کنترل تصادفی بودن.....
۲۱۱	۵-۴ بارگذاری و ذخیره‌سازی وزن‌های مدل در PyTorch.....
۲۱۳	۵-۵ بارگذاری وزن‌های از پیش آموزش داده‌شده از OpenAI.....
۲۲۲	خلاصه.....

فصل ششم: تنظیم دقیق برای طبقه‌بندی..... ۲۲۳

۲۲۴	۶-۱ روش‌های مختلف تنظیم دقیق.....
۲۲۶	۶-۲ آماده‌سازی مجموعه داده.....
۲۳۱	۶-۳ ایجاد بارگذارهای داده.....
۲۳۷	۶-۴ مقداردهی اولیه مدل با وزن‌های از پیش آموزش داده شده.....
۲۴۰	۶-۵ اضافه کردن یک سر طبقه‌بندی.....
۲۴۸	۶-۶ محاسبه‌ی خطا و درستی طبقه‌بندی.....
۲۵۳	۶-۷ تنظیم دقیق مدل روی داده‌های نظارت‌شده.....
۲۶۰	۶-۸ استفاده از مدل زبانی بزرگ به‌عنوان یک طبقه‌بند هرزنامه.....
۲۶۲	خلاصه.....

فصل هفتم: تنظیم دقیق برای پیروی کردن از دستورعمل‌ها..... ۲۶۵

۲۶۶	۷-۱ مقدمه‌ای بر تنظیم دقیق دستورعمل.....
۲۶۷	۷-۲ آماده‌سازی مجموعه داده برای تنظیم دقیق دستورعمل نظارت‌شده.....
۲۷۳	۷-۳ سازماندهی داده‌ها در دسته‌های آموزشی.....
۲۸۷	۷-۴ ایجاد بارگذارهای داده برای یک مجموعه داده دستورعمل.....
۲۹۰	۷-۵ بارگذاری یک مدل زبانی بزرگ از پیش آموزش دیده.....
۲۹۴	۷-۶ بهینه‌سازی مدل زبانی بزرگ روی داده‌های دستورعمل.....
۲۹۹	۷-۷ استخراج و ذخیره پاسخ‌ها.....
۳۰۵	۷-۸ ارزیابی مدل زبانی بزرگ به‌دقت تنظیم‌شده.....
۳۱۶	۷-۹ نتیجه‌گیری.....
۳۱۸	خلاصه.....

پیوست الف: مقدمه‌ای بر PyTorch	۳۱۹
الف-۱ PyTorch چیست؟	۳۱۹
الف-۲ درک تنسورها	۳۲۷
الف-۳ مدل‌ها به‌عنوان نمودارهای محاسباتی	۳۳۲
الف-۸ مشتق‌گیری خودکار آسان می‌شود	۳۳۳
الف-۵ پیاده‌سازی شبکه‌های عصبی چندلایه	۳۳۶
الف-۶ راه‌اندازی بارگذارهای داده کارآمد	۳۴۲
الف-۷ یک حلقه آموزشی معمولی	۳۴۸
الف-۸ ذخیره و بارگذاری مدل‌ها	۳۵۳
الف-۹ بهینه‌سازی عملکرد آموزش با GPUها	۳۵۴
خلاصه	۳۶۵

پیوست ب: منابع و مطالعه بیشتر

پیوست د: افزودن ویژگی‌های تکمیلی به حلقه آموزش	۳۹۷
د-۱ نرخ یادگیری	۳۹۹
د-۲ کاهش کسینوسی	۴۰۱
د-۳ برش گرادیان	۴۰۲
د-۴ تابع آموزشی اصلاح‌شده	۴۰۴

پیوست ه: تنظیم دقیق با بهره‌وری پارامتر با استفاده از LoRA

ه-۱ مقدمه‌ای بر LoRA	۴۰۹
ه-۲ آماده‌سازی مجموعه‌داده	۴۱۱
ه-۳ مقداردهی اولیه مدل	۴۱۴
ه-۴ تنظیم دقیق با کارایی پارامتر بالا با استفاده از LoRA	۴۱۶

پیشگفتار

من همیشه شیفته‌ی مدل‌های زبانی بوده‌ام. بیش از یک دهه پیش، سفر من به دنیای هوش مصنوعی با یک کلاس طبقه‌بندی الگوهای آماری آغاز شد که منجر به اولین پروژه‌ی مستقل من شد: توسعه‌ی یک مدل و برنامه‌ی کاربردی وب برای تشخیص حالت یک آهنگ بر اساس متن ترانه‌ی آن. با گذشت زمان و رسیدن به سال ۲۰۲۲، با انتشار ChatGPT، مدل‌های زبانی بزرگ^۱ (LLMs) جهان را تحت تأثیر قرار دادند و انقلابی در نحوه‌ی کار بسیاری از ما ایجاد کردند. این مدل‌ها به شکل بی‌نظیری همه‌کاره هستند و در وظایفی مانند بررسی گرامر، نوشتن ایمیل، خلاصه‌کردن اسناد طولانی و موارد دیگر کمک می‌کنند. این امر مدیون توانایی آن‌ها در تجزیه و تولید متن شبیه به انسان است که در زمینه‌های مختلف، از خدمات مشتری گرفته تا تولید محتوا و حتی در حوزه‌های فنی‌تر مانند کدنویسی و تحلیل داده، اهمیت دارد.

همان‌گونه که از نامشان پیداست، یکی از ویژگی‌های بارز مدل‌های زبانی بزرگ این است که بزرگ هستند- بسیار بزرگ- و شامل میلیون‌ها تا میلیارد‌ها پارامتر می‌شوند (برای مقایسه، با استفاده از روش‌های سنتی‌تر یادگیری ماشین یا آماری، مجموعه‌داده‌ی Iris را می‌توان با استفاده از یک مدل کوچک با تنها دو پارامتر با درستی^۲ بیش از ۹۰٪ طبقه‌بندی کرد). با این حال، علی‌رغم اندازه‌ی بزرگ مدل‌های زبانی بزرگ در مقایسه با روش‌های سنتی‌تر، مدل‌های زبانی بزرگ مجبور نیستند یک جعبه‌ی سیاه باشند.

در این کتاب، شما یاد خواهید گرفت که چگونه یک مدل زبانی بزرگ را گام به گام بسازید. در پایان، شما درک کاملی از نحوه کار یک مدل زبانی بزرگ، مانند مواردی که در ChatGPT استفاده می‌شوند، در سطح بنیادی خواهید داشت. من معتقدم که ایجاد اعتماد به هر بخش از مفاهیم اساسی و کد مرتبط برای موفقیت بسیار مهم است. این امر نه تنها به رفع اشکالات و بهبود عملکرد کمک می‌کند، بلکه امکان آزمایش ایده‌های جدید را نیز فراهم می‌سازد.

چند سال پیش، زمانی که شروع به کار با مدل‌های زبانی بزرگ کردم، مجبور شدم نحوه‌ی پیاده‌سازی آن‌ها را به‌سختی یاد بگیرم و برای رسیدن به یک درک کلی، مقالات تحقیقاتی و مخازن کد ناقص زیادی را بررسی کنم. با این کتاب، امیدوارم با توسعه و به اشتراک گذاشتن یک آموزش پیاده‌سازی گام به گام که تمام اجزای اصلی و فازهای توسعه‌ی یک مدل زبانی بزرگ را به تفصیل شرح می‌دهد، مدل‌های زبانی بزرگ را در دسترس‌تر کنم.

1- Large Language Models

2- Accuracy

من کاملاً معتقدم که بهترین راه برای درک مدل‌های زبانی بزرگ، کدنویسی یکی از آنها از ابتدا است و خواهید دید که این امر می‌تواند سرگرم‌کننده نیز باشد!

درباره این کتاب

این کتاب نوشته شده است تا به شما در درک و ساخت مدل‌های زبانی بزرگ شبیه به GPT از پایه کمک کند. این کتاب با تمرکز بر اصول کار با داده‌های متنی و کدنویسی سازوکارهای توجه^۱ آغاز می‌شود و سپس شما را در پیاده‌سازی یک مدل کامل GPT از ابتدا راهنمایی می‌کند. سپس سازوکار پیش‌آموزش^۲ و هم‌چنین تنظیم دقیق^۳ برای وظایف خاص مانند طبقه‌بندی متن و پیروی از دستورعمل‌ها بیان می‌شود. در پایان این کتاب، شما درک عمیقی از نحوه کار مدل‌های زبانی بزرگ و مهارت‌های لازم برای ساخت مدل‌های خودتان خواهید داشت. مدل‌هایی که شما خواهید ساخت در مقایسه با مدل‌های پایه بزرگ، در مقیاس کوچک‌تری هستند، اما از مفاهیم مشابهی استفاده می‌کنند و به‌عنوان ابزارهای آموزشی قدرتمندی برای درک سازوکارها و تکنیک‌های اصلی مورد استفاده در ساخت مدل‌های زبانی بزرگ پیشرفته عمل می‌کنند.

1- Attention Mechanisms
2- Pretraining
3- Fine-Tuning