



درس هوش مصنوعی : پروژه نهایی

استاد: خانم دکتر بشری پیشگو

دستیار ارشد: پژمان زیوری

دستیاران طراح: سمانه تائبی، امیر حسین حنیفه پور، پژمان زیوری

عنوان پروژه : تولید کپشن تصویر با استفاده از مدل‌های CNN و LSTM

مقدمه

یکی از چالش‌های جذاب و کاربردی در حوزه هوش مصنوعی، توانایی کامپیوترها در درک و توصیف تصاویر است. پروژه تولید کپشن تصویر فرصتی فراهم می‌آورد تا دانشجویان با تکنیک‌های پیشرفته یادگیری عمیق آشنا شوند. هدف اصلی این پروژه، طراحی سیستمی است که بتواند با ترکیب شبکه‌های عصبی پیچشی (Convolutional Neural Networks - CNN) برای استخراج ویژگی‌های بصری و شبکه‌های عصبی حافظه‌ی طولانی کوتاه‌مدت (Long Short-Term Memory - LSTM) برای تولید توالی‌های متنی، توضیحاتی دقیق برای تصاویر ارائه دهد.

اهداف پروژه

- آشنایی با اصول و کاربردهای شبکه‌های عصبی پیچشی (CNN) در استخراج ویژگی‌های تصاویر.
- یادگیری نحوه استفاده از شبکه‌های عصبی بازگشتی (Recurrent Neural Networks - RNN) به‌ویژه شبکه عصبی حافظه‌ی طولانی کوتاه‌مدت در تولید توالی‌های متنی.
- تقویت مهارت‌های مربوط به پردازش و آماده‌سازی داده‌ها برای مدل‌های یادگیری عمیق.
- پیاده‌سازی یک سیستم انتها به انتها (End-to-End) برای تولید توضیحات متنی مرتبط با تصاویر.
- بهبود توانایی تحلیل عملکرد مدل‌ها و ارائه راهکارهای بهینه‌سازی.

دیتاست مورد استفاده

Flickr_8K Dataset

دیتاست مورد استفاده شامل:

- ۸۰۰۰ تصویر متنوع از محیط‌های واقعی.
- ۵ کپشن متنی برای هر تصویر که توسط افراد مختلف نوشته شده است.
- دانشجویان باید دیتاست را دانلود و آماده‌سازی کنند. این آماده‌سازی شامل:
- جداسازی داده‌ها برای آموزش، اعتبارسنجی و آزمون.
- پردازش تصاویر (تغییر اندازه و نرمال‌سازی).
- پردازش متن‌ها (توکنایز کردن، ساخت واژه‌نامه و تبدیل متن به قالب عددی).

مدل‌ها و تکنیک‌های مورد استفاده

۱. شبکه عصبی پیچشی (CNN):

در این پروژه، از مدل پیش‌آموزش‌دیده **Xception** برای استخراج ویژگی‌های بصری استفاده می‌شود. در این فرآیند:

- لایه‌های انتهایی مدل که برای دسته‌بندی طراحی شده‌اند، حذف می‌شوند.
 - خروجی لایه‌های میانی به‌عنوان ویژگی‌های استخراج‌شده تصویر در نظر گرفته می‌شود.
- این ویژگی‌ها، نمایشی غنی و فشرده از اطلاعات بصری تصویر هستند و برای تولید کپشن به مدل شبکه عصبی حافظه‌ی طولانی کوتاه‌مدت منتقل می‌شوند.

۲. شبکه عصبی حافظه‌ی طولانی کوتاه‌مدت:

یک مدل طراحی می‌شود که بتواند:

- بر اساس ویژگی‌های استخراج‌شده از تصویر و کلمات قبلی، کلمه بعدی را پیش‌بینی کند.
- برای مدیریت کلمات، یک واژه‌نامه ایجاد می‌شود که کلمات پرتکرار را شامل شده و کلمات کم‌تکرار را به‌صورت "UNK" ناشناخته (علامت‌گذاری می‌کند).

این فرآیند به مدل کمک می‌کند تا روی کلمات مهم تمرکز کند و حجم محاسباتی کاهش یابد.

۳. لایه تعبیه (Embedding Layer):

لایه‌های تعبیه برای نمایش برداری کلمات به کار می‌روند تا مقادیر عددی قابل فهم برای مدل ایجاد کنند.

۴. تکنیک‌های بهبود:

- استفاده از روش‌هایی مانند **Early Stopping** برای جلوگیری از بیش‌برازش (**Overfitting**).
- اعمال **Dropout** برای افزایش تعمیم‌پذیری مدل.
- تنظیم نرخ یادگیری (**Learning Rate**) برای بهبود فرآیند آموزش و همگرایی سریع‌تر.

۵. ترکیب مدل‌ها و آموزش نهایی:

- مدل شبکه عصبی پیچشی و شبکه عصبی حافظه‌ی طولانی کوتاه‌مدت ترکیب شده و یک سیستم انتها به انتها (**End-to-End**) برای تولید کپشن ایجاد می‌شود.
- مدل با داده‌های آموزشی تمرین داده می‌شود تا توالی‌های متنی مرتبط با تصاویر تولید کند.
- تکنیک‌های بهینه‌سازی برای بهبود دقت و پایداری مدل اعمال می‌شود.

ارزیابی و نمایش نتایج

۱. ارزیابی مدل:

- مدل با استفاده از داده‌های تست ارزیابی می‌شود.
- از معیارهای استاندارد مانند **BLEU Score**، **METEOR** و **CIDEr** برای سنجش کیفیت کپشن‌های تولیدشده استفاده می‌شود.
- این معیارها امکان مقایسه کپشن‌های تولیدشده با کپشن‌های واقعی را فراهم می‌کنند و نقاط قوت و ضعف مدل را آشکار می‌سازند.

۲. نمایش نتایج:

- نمونه‌هایی از تصاویر همراه با کپشن‌های واقعی و کپشن‌های تولیدشده توسط مدل ارائه می‌شود.
- این نمایش می‌تواند به صورت جداول یا گرافیک‌های بصری باشد تا عملکرد مدل به وضوح مشخص شود.
- تحلیل کنید که مدل در چه مواردی عملکرد خوبی داشته است و در چه مواردی نیاز به بهبود دارد.
- مثال‌هایی از خطاهای مدل و دلایل احتمالی آن‌ها را ارائه دهید.

۳. تهیه گزارش نهایی:

گزارشی شامل موارد زیر تهیه کنید:

- مقدمه: شرح مسئله و اهمیت آن.
- داده‌ها و پیش‌پردازش: توضیح فرآیند آماده‌سازی داده‌ها.
- مدل و معماری: شرح معماری مدل و فرآیند ترکیب شبکه عصبی پیچشی و شبکه عصبی حافظه‌ی طولانی کوتاه‌مدت.
- نتایج و ارزیابی: ارائه نتایج آموزشی، ارزیابی مدل و تحلیل آن.
- نتیجه‌گیری و پیشنهادات: تحلیل عملکرد مدل و ارائه راهکارهایی برای بهبود در آینده.

- **کد پروژه:** ضمیمه کدهای پروژه به صورت مستند یا آپلود آن‌ها در یک مخزن **GitHub**.

نکات تکمیلی

مدیریت منابع محاسباتی:

- با توجه به پیچیدگی مدل و حجم داده‌ها، استفاده از **GPU** برای تسریع فرآیند آموزش پیشنهاد می‌شود.

مستندسازی کد:

- کدها را با توضیحات مناسب مستندسازی کنید تا قابلیت درک و بازبینی داشته باشند.

استفاده از مراجع:

- منابع استفاده‌شده یا ایده‌های گرفته‌شده را در گزارش خود ذکر کنید تا اعتبار پروژه تضمین شود.

اصول اخلاقی:

- اطمینان حاصل کنید که استفاده از داده‌ها و مدل‌ها مطابق با قوانین حق نشر و اخلاق پژوهشی باشد.

نکات امتیازی

برای دریافت نمره امتیازی، می‌توانید:

- به جای مدل **Xception**، از یکی از معماری‌های جایگزین مانند **VGG16**، **InceptionV3** یا **ResNet50** برای استخراج ویژگی‌های بصری استفاده کنید.
- نتایج به دست آمده را با مدل اصلی مقایسه کرده و تأثیر معماری‌های مختلف بر کیفیت کپشن‌های تولیدی را ارزیابی کنید.
- گزارش مختصری ارائه دهید که شامل تحلیل عملکرد مدل‌ها باشد تا درک معماری‌های مختلف و تأثیر آن‌ها بر پروژه افزایش یابد.

مهلت تحویل

پروژه باید تا تاریخ تعیین شده توسط استاد، به همراه تمامی ضمائم تحویل داده شود.