دانشگاه تهران

دانشکده علوم و فنون نوین

گزارش هفته دوم و سوم

نام و نام خانوادگی	فاطمه چیت ساز
شماره دانشجویی	830402092
تاریخ ارسال گزارش	6 دی 1402

سرتیتر ها

هفته دوم : مطالعه بر روی یک سری چیز های base

در این هفته من سعی کردم یک سری مطالب مهم که base خیلی الگوریتم هاست رو بخونم:

۱. شبکه های عصبی بازگشتی (RNN) و پیش بینی سهام سه روزه:

RNN ها شبکه هایی هستن که اطلاعات گذشته رو به خاطر می سپارن و برای کارهای مثل پیش بینی سهام مفیده.

یه مشکل این شبکه ها اینه که اگه وزن ها خیلی بزرگ باشن (مثلا ۲)، اطلاعات گذشته خیلی زیاد تو وزن ها تاثیر می ذاره و نتیجه اشتباه میشه (مثل انفجار).

اگه وزن ها خیلی کوچیک باشن (مثلا ۰.۵)، اطلاعات گذشته خیلی کم تو وزن ها تاثیر می ذاره و دیگه چیزی یاد نمیگیره (مثل محو شدن).

۲. شبکه های عصبی حافظه بلندمدت (LSTM):

LSTM ها یه نوع RNN پیشرفته هستن که دو تا مسیر دارن: یکی برای حافظه کوتاه مدت و یکی برای حافظه بلندمدت.

حافظه بلندمدت اطلاعات مهم رو نگه می داره و حافظه کوتاه مدت اطلاعات جدید رو.

LSTM ها از یه سری تابع های ریاضی استفاده می کنن (مثل سیگموید و tanh) تا مشخص کنن چه مقدار از اطلاعات گذشته رو نگه دارن.

با این کار، LSTM ها می تونن برای کارهای مثل پیش بینی سهام سه روزه که به اطلاعات گذشته نیاز داره، بهتر از RNN ها عمل کنن.

۳. word2vec:

شبکه های عصبی با کلمات خوب کار نمی کنن چون کلمات حروف و اعداد هستن که معنای خاصی ندارن.

word2vec یه الگوریتمیه که به هر کلمه یه عدد خاص میده و این عددها طوری انتخاب میشن که کلمات مشابه، اعداد نزدیک به هم داشته باشن.

word2vec می تونه به شبکه های عصبی کمک کنه تا بفهمه کلمات چطوری با هم مرتبط هستن و مثلا کلمه بعدی یه جمله رو پیش بینی کنه.

۴. مدل رمزگذار-رمزگشا (seq2seq):

این مدل برای ترجمه زبان ها یا تبدیل متن به صوت استفاده میشه.

بخش رمزگذار، متن ورودی رو به یه بردار خلاصه (context vector) تبدیل می کنه.

بخش رمزگشا، از این بردار خلاصه و کلمه قبلی استفاده می کنه تا کلمه بعدی رو پیش بینی کنه.

این مدل برای ترجمه زبان هایی که طول جملاتشون متفاوته، چالش برانگیزه.

۵. توجه (attention):

توجه یه مکانیسمه که به مدل کمک می کنه تا روی بخش های مهم ورودی تمرکز کنه.

مثلا، برای ترجمه جمله "don't eat the apple" مدل باید روی کلمه "don't" بیشتر تمرکز کنه تا جمله رو درست ترجمه کنه.

توجه با محاسبه شباهت بین کلمات و استفاده از تابع سافت‌مکس کار می کنه.

۶. شبکه های عصبی ترانسفورماتور:

شبکه‌های عصبی Transformer برای درک ارتباط بین کلمات در جملات و متون طراحی شده‌اند. این شبکه‌ها به ویژه در زمینه ترجمه ماشینی، پاسخگویی به سؤالات و خلاصه‌سازی متن بسیار کارآمد هستند.

اجزای اصلی شبکه Transformer عبارتند از:

(Positional Encoding):

برای حفظ ترتیب کلمات در جمله به هر کلمه یک بردار عددی اختصاص می‌دهد.

مثال: در جمله "من بستنی دوست دارم"، کدگذاری موقعیتی به کلمه "من" یک بردار متفاوت از کلمه "بستنی" می‌دهد تا نشان دهد که این کلمات در جایگاه‌های متفاوتی از جمله قرار دارند.

(Self-Attention):

به هر کلمه اجازه می‌دهد به سایر کلمات در جمله توجه کند تا ارتباط بین آن‌ها را پیدا کند.

برای هر کلمه، سه بردار تولید می‌کند: (Query)، (Key) و (Value)

برای محاسبه میزان شباهت هر کلمه با سایر کلمات، از ضرب نقطه‌ای بین بردارهای Query و کلید آن‌ها استفاده می‌شود.

سپس، از تابع سافت‌مکس برای نرمال‌سازی این شباهت‌ها و تبدیل آن‌ها به احتمال استفاده می‌شود.

در نهایت، بردار مقدار هر کلمه با احتمال مربوط به آن وزن‌دهی شده و با هم جمع می‌شوند تا بردار توجه به خود برای آن کلمه تولید شود.

مثال: در جمله "من پیتزای آن را خوردم"، توجه به خود به مدل کمک می‌کند تا درک کند که کلمه "آن" به "پیتزا" اشاره دارد.

(Residual Connections):

برای تسهیل یادگیری و جلوگیری از از بین رفتن اطلاعات در لایه‌های عمیق شبکه استفاده می‌شوند.

این اتصالات اجازه می‌دهند اطلاعات از لایه‌های قبلی مستقیماً به لایه‌های بعدی منتقل شوند.

رمزگذار-رمزگشا (Encoder-Decoder):

از دو بخش اصلی تشکیل شده است: رمزگذار و رمزگشا

رمزگذار: ورودی را پردازش می‌کند و یک نمایش برداری از آن تولید می‌کند.

رمزگشا: از این نمایش برداری برای تولید خروجی استفاده می‌کند.

در ترجمه ماشینی، رمزگذار جمله اصلی را پردازش می‌کند و رمزگشا جمله ترجمه شده را تولید می‌کند.

توجه رمزگذار-رمزگشا (Encoder-Decoder Attention):

به رمزگشا اجازه می‌دهد به بخش‌های مرتبط در ورودی اصلی توجه کند.

این کار با استفاده از بردارهای کوئری و کلید مشابهی که در توجه به خود استفاده شد، انجام می‌شود.

شبکه کاملاً متصل (Fully Connected Network):

در نهایت، خروجی رمزگشا به یک شبکه کاملاً متصل وارد می‌شود که احتمال هر کلمه در خروجی را تولید می‌کند.

تابع سافت‌مکس برای نرمال‌سازی این احتمالات و تعیین کلمه‌ای که باید انتخاب شود، استفاده می‌شود.۷.
شبکه های عصبی آلیس:

شبکه های آلیس نوعی RNN هستن که از توجه استفاده می کنن.

این شبکه ها برای کارهایی مثل تشخیص گفتار و تولید موسیقی استفاده میشن.

مجموعه ویدیو مناسب برای برسی بیشتر :

https://www.youtube.com/watch?v=zxagGtF9MeU&list=PLblh5JKOoLUIxGDQs4LFFD--41Vzf-ME1

هفته سوم : برسی اینکه دعوا سر چیه

۱. Weak-to-strong generalization: رویکردی امیدوارکننده برای تطبیق هوش مصنوعی

این مقاله ایده جالبی بهمون میده: چطور یک مدل هوش مصنوعی کوچکتر و "معلم"تر، می تونه مدل بزرگتر و "شاگرد" رو راهنمایی کنه تا عملکرد ایمن‌تر و مفیدتری داشته باشه.

تصور کنید یه مدل کوچیک و باتجربه داریم که می‌دونه چطور تو دنیای واقعی رفتار کنه و از خطرات و اشتباهات اجتناب کنه. حالا فکر کنید این مدل بخواد دانشش رو به یه مدل بزرگ‌تر و قوی‌تر منتقل کنه تا اون هم بتونه درست و ایمن عمل کنه. این دقیقا کاریه که "Weak-to-strong generalization" انجام می ده.

این روش چند تا مزیت مهم داره:

هوش مصنوعی فراتر از هوش ما: مدل "شاگرد" می تونه با کمک مدل "معلم" حتی از خودش هوشمندتر بشه و کارهای پیچیده‌تری انجام بده.

ایمنی بیشتر: مدل "معلم" می‌تونه به مدل "شاگرد" بگه چطور از خطرات و رفتارهای نادرست اجتناب کنه و خطرات رو پیش‌بینی کنه.

کاربردی‌تر شدن هوش مصنوعی: مدل "شاگرد" با آموزش‌های مدل "معلم" می‌تونه وظایف مفیدتری رو انجام بده و به حل مشکلات واقعی کمک کنه.

البته این ایده هنوز چالش‌های خودش رو داره:

چطور مدل "معلم" رو باید آموزش بدیم؟ آموزش مدل "معلم" به اندازه کافی سخت هست که اون خودش باید فهم عمیقی از نحوه‌ی رفتار ایمن و مفید داشته باشه.

چطور مطمئن بشیم مدل "معلم" قابل اعتماده؟ اگر مدل "معلم" قابل اعتماد نباشه، ممکنه دانش اشتباه یا خطرناک به مدل "شاگرد" منتقل کنه.

چطور حواس‌مون باشه مدل "شاگرد" حرف "معلم" رو گوش می‌کنه؟ ممکنه مدل "شاگرد" باهوش‌تر از مدل "معلم" بشه و دیگه از راهنمایی‌های اون پیروی نکنه.

در کل، Weak-to-strong generalization یه نوید بزرگ برای تطبیق هوش مصنوعی داره. اما تحقیقات بیشتری لازمه تا بتونیم به طور کامل از پتانسیل اون استفاده کنیم و مطمئن بشیم که هوش مصنوعی در مسیر درستی حرکت می‌کنه.

https://openai.com/research/weak-to-strong-generalization

۲. Practices for Governing Agentic AI Systems: چطور هوش مصنوعی مستقل رو کنترل کنیم؟

مقاله دوم روی جنبه دیگری از تطبیق هوش مصنوعی تمرکز می کنه: "هوش مصنوعی مستقل" یا Agentic AI. این سیستم‌های هوش مصنوعی می‌تونن به صورت مستقل عمل کنن و تصمیم بگیرن. یه ربات جراح که خودش عمل جراحی انجام می‌ده، یه مثال از هوش مصنوعی مستقل هست.

این سیستم‌ها مزیت‌های زیادی دارن و می‌تونن مشکلات پیچیده رو حل کنن و زندگی انسان‌ها رو بهبود ببخشن. اما در عین حال، خطراتی هم به همراه دارن. اگر کنترل این سیستم‌ها رو از دست بدیم، ممکنه عواقب جبران‌ناپذیری داشته باشه.

برای اطمینان از توسعه و استفاده‌ی ایمن و مسئولانه‌ی هوش مصنوعی مستقل، به قوانین و مقررات مناسب نیاز داریم. این قوانین باید:

شفاف: مردم باید بدونن هوش مصنوعی چطور کار می‌کنه و چه تصمیم‌هایی می‌گیره.

پاسخگو: باید مشخص باشه چه کسی مسئول عملکرد هوش مصنوعی مستقل هست.

ایمن: باید تمهیداتی برای جلوگیری از خطرات بالقوه‌ی این سیستم‌ها در نظر گرفته بشه.

هماهنگ با ارزش‌های انسانی: هوش مصنوعی باید طوری طراحی بشه که با ارزش‌های انسانی مثل عدالت، انصاف و شفقت سازگار باشه.

اخلاقی: هوش مصنوعی باید طوری عمل کنه که از نظر اخلاقی قابل قبول باشه.

لینک:

https://openai.com/research/practices-for-governing-agentic-ai-systems

3. DALL-E 3

VQ-VAE: قطعات کد برای ساخت تصویر

DALL-E 3 از یه تکنولوژی به نام VQ-VAE استفاده می‌کنه. تصور کنید یه جعبه ابزار با هزاران قطعه‌ی کد دارید. VQ-VAE متن توصیفی شما رو به این قطعات کد تبدیل می‌کنه و بعد از اون‌ها برای ساخت تصویر استفاده می‌کنه. هر کد یه بخش کوچیک از تصویر رو مشخص می‌کنه، مثل رنگ یه نقطه یا شکل یه خط. بعد، DALL-E 3 این قطعه‌ها رو کنار هم می‌ذاره تا تصویر نهایی رو بسازه.

ChatGPT: مترجم متن به تصویر

گاهی وقت‌ها توصیف‌های ما خیلی مبهمه، مثلا می‌گیم "یه گربه بامزه که تو یه مزرعه بازی می‌کنه." DALL-E 3 ممکنه گیج بشه که چه شکلی دقیقا می‌خوایم. اینجا ChatGPT به کمکش میاد! ChatGPT می‌تونه توصیف کوتاه ما رو به یه جمله‌ی دقیق‌تر و جزئی‌تر تبدیل کنه، مثلا "یه گربهٔ نارنجی با چشم‌های سبز که تو یه مزرعه پر از گل‌های آفتاب‌گردان داره با یه پروانه بازی می‌کنه." با این توصیف دقیق‌تر، DALL-E 3 می‌تونه تصویر دقیق‌تری بسازه.

چالش‌های DALL-E 3:

DALL-E 3 هنوز یه سیستم جدیده و در حال یادگیریه، پس طبیعیه که یه سری چالش هم داشته باشه:

کیفیت تصویر: DALL-E 3 با عکس‌های خیلی باکیفیت و جزئیات زیاد مشکل داره. ممکنه عکس‌هایی که می‌سازه کمی تار یا مصنوعی به نظر برسن.

آدم‌های واقعی: ساختن تصویر آدم‌های واقعی، خصوصا صورتشون، یه چالش بزرگه. ممکنه عکس‌ها غیرطبیعی یا حتی توهین‌آمیز باشن.

مفاهیم پیچیده: تبدیل ایده‌های انتزاعی یا ظریف به تصویر می‌تونه سخت باشه. ممکنه عکس‌هایی که می‌سازه مبهم یا گیج‌کننده باشن.

سوگیری: DALL-E 3 هم مثل بقیه‌ی سیستم‌های هوش مصنوعی، ممکنه سوگیری‌های موجود تو اطلاعاتی که باهاش آموزش دیده رو به ارث ببره و عکس‌های غیرمنصفانه یا تبعیض‌آمیز بسازه.

کنترل خلاقانه: هنوز سخته که دقیقا همون تصویری رو که تو ذهنتونه بسازیم. ممکنه نتیجه‌ی نهایی با چیزی که می‌خواستیم خیلی فرق داشته باشه.

دسترسی‌پذیری: فعلا فقط تعداد محدودی از افراد به DALL-E 3 دسترسی دارن. این یه نگرانی ایجاد می‌کنه که نکنه خلاقیت دست یه عده‌ی خاص بمونه و همه‌گیر نشه.

هزینه: هنوز مشخص نیست که استفاده از DALL-E 3 چقدر هزینه داشته باشه. این ممکنه باعث بشه خیلی‌ها نتونن ازش استفاده کنن.

ملاحظات اخلاقی: امکان ساختن عکس‌های واقعی‌نما ولی جعلی، نگرانی‌هایی رو درباره اطلاعات غلط، deepfakes و مالکیت آثار دیجیتال ایجاد می‌کنه.

مقاله :

https://cdn.openai.com/papers/dall-e-3.pdf

4. GPT-4v: زبان‌های بزرگ سریع‌تر یاد می‌گیرند!

تصور کنید به یک مدل زبان بگویید یک شعر عاشقانه بنویسد، یک ایمیل رسمی برای رئیس‌تان آماده کند، یا حتی کد یک بازی ساده را تولید کند. GPT-4v جدیدترین مدل زبان از OpenAI است که می‌تواند با حداقل داده‌ی آموزشی، انواع کارهای مختلف را انجام دهد. این مدل باهوش، به سرعت یاد می‌گیرد و می‌تواند خروجی‌های باکیفیتی تولید کند.

اما داستان به همین سادگی تمام نمی‌شود. هر تکنولوژی قدرتمندی، چالش‌هایی هم دارد. یکی از نگرانی‌های اصلی درباره GPT-4v، سوگیری (bias) است. ممکن است داده‌هایی که مدل با آن‌ها آموزش دیده، به طور ناخواسته باعث ایجاد تعصباتی در خروجی‌های آن شود. برای مثال، اگر مدل بیشتر با متون ادبیات مردانه آموزش دیده باشد، ممکن است خروجی‌های آن در مورد زنان یا موضوعات مرتبط با زنان، دچار سوگیری شود.

چالش دیگر، درک نحوه‌ی تصمیم‌گیری و محاسبات داخلی GPT-4v است. این مدل پیچیده، مانند یک جعبه سیاه عمل می‌کند و فهمیدن اینکه چطور به نتایج خاصی می‌رسد، برای ما دشوار است. این موضوع کنترل دقیق عملکرد مدل را به چالش می‌کشد.

5. استدلال‌های ریاضی:

یکی دیگر از خبرهای خوب ، پیشرفت قابل توجه در توانایی استدلال ریاضی زبان‌های بزرگ است. تا پیش از این، مدل‌های زبان بزرگ در حل مسائل ریاضی عملکرد چندان خوبی نداشتند. اما محققان OpenAI با روشی به نام "راهنمایی فرآیندی" (process supervision) توانستند عملکرد GPT-3 در حل مسائل ریاضی را به شکل چشمگیری بهبود ببخشند.

در این روش، به مدل علاوه بر مسئله‌ی اصلی، مراحل و استدلال‌های حل آن هم آموزش داده می‌شود. این کار مانند آن است که یک معلم ریاضی به جای فقط دادن جواب نهایی، مراحل حل مسئله را هم به شاگردان توضیح دهد. نتایج آزمایش‌ها نشان می‌دهد که با استفاده از این روش، GPT-3 توانسته طیف وسیعی از مسائل ریاضی را با دقت بسیار بالایی حل کند.

البته، این روش هم بی‌عیب نیست. برای اجرای راهنمایی فرآیندی، وجود پایگاه داده‌ای عظیم از مسائل ریاضی همراه با مراحل حل آن‌ها ضروری است. جمع‌آوری و برچسب‌گذاری چنین داده‌هایی امری زمان‌بر و پرهزینه است. علاوه بر این، استفاده از این روش می‌تواند پیچیدگی مدل را افزایش دهد و کنترل آن را دشوارتر کند.

https://openai.com/research/improving-mathematical-reasoning-with-process-supervision

6. خودتوضیحی در مدل‌های زبانی:

https://openai.com/research/language-models-can-explain-neurons-in-language-models

محققان OpenAI روشی برای استفاده از مدل‌های زبانی بزرگ (LLM) برای توضیح رفتار مدل‌های زبانی کوچک‌تر ارائه کردند.

در این پژوهش، از GPT-3 برای تولید توضیحات متنی درباره‌ی فعالیت نورون‌ها در مدل GLUE (که برای انجام وظایف پردازش زبان طبیعی آموزش دیده) استفاده شد.

نتایج نشان داد که GPT-3 توانسته است توضیحات معنادار و مرتبط با مفاهیم زبانی مانند کلمات، عبارات و جملات را برای فعالیت نورون‌ها ارائه دهد.

محدودیت این پژوهش، استفاده از تنها یک LLM (GPT-3) و یک مجموعه داده (GLUE) بود.

7. GPTها و تأثیرات بالقوه بر بازار کار:

این پژوهش به بررسی پتانسیل GPTها برای ایجاد تغییرات گسترده در بازار کار می‌پردازد.

هدف اصلی، ایجاد یک GPT عمومی و در دسترس برای همه است که می‌تواند در هر زمینه‌ای تخصص داشته باشد.

این پژوهش به بررسی تأثیرات بالقوه‌ی GPTها بر مشاغل مختلف و نیاز به مهارت‌های جدید در نیروی کار می‌پردازد.

8. مدل‌های زبانی، یادگیرندگان سریع

این مقاله نشان می‌دهد که مدل‌های زبانی بزرگ مانند GPT-3 می‌توانند با حداقل مثال یا دستورالعمل، عملکردی چشمگیر در وظایف مختلف داشته باشند.

این ویژگی در تضاد با روش‌های سنتی یادگیری ماشین است که برای هر وظیفه نیاز به تنظیم دقیق و آموزش فراوان دارند.

یادگیری سریع می‌تواند کاربردهای گسترده‌ای در زمینه‌های مختلف مانند ترجمه‌ی زبان، تولید متن، خلاصه‌نویسی و بسیاری از وظایف دیگر داشته باشد.

https://arxiv.org/abs/2005.14165

9. contrails

/https://sites.research.google/contrails

هواپیماها با ردی از ابرهای خطی به نام «کنترا» (Contrail) آسمان را طی می‌کنند. این ابرها با به دام انداختن گرما، در گرم شدن زمین نقش دارند و نوع شبانه‌ی آن‌ها اثری بیشتر دارد. محققان Google AI با استفاده از هوش مصنوعی و تحلیل حجم عظیمی از داده‌های هواشناسی، ماهواره و پرواز، روشی برای پیش‌بینی و اجتناب از ایجاد این ابرها ارائه کرده‌اند. این همکاری با خطوط هوایی American Airlines منجر به کاهش قابل توجه ۵۴ درصدی تولید کنترا شد، با این حال پروازهایی که برای اجتناب از کنترا مسیر خود را تغییر دادند، ۲ درصد سوخت بیشتری مصرف کردند. هدف از این روش، ارائه‌ی راهکاری مقرون‌به‌صرفه برای مقابله با تغییرات اقلیمی است.

10. ML Kit در اندروید

Google AI با انتشار ML Kit، پلتفرمی جدید برای هوش مصنوعی در این سیستم‌عامل، امکانات جالبی را به گوشی‌های شما می‌آورد. یکی از نمونه‌ها، استفاده از «پاسخ هوشمند» (Smart Reply) در

کیبورد درون واتساپ با استفاده از مدل Gemini Nano در گوشی Pixel 8 Pro است. این قابلیت با درک متن پیام‌های دریافتی، پاسخ‌های کوتاه و مناسبی پیشنهاد می‌کند.

علاوه بر این، ML Kit با رویکردی به نام LoRA (Low Rank Adaptation) به توسعه‌دهندگان اپلیکیشن‌ها امکان می‌دهد مدل‌های زبان بزرگ را برای نیازهای خاص خود سفارشی کنند. با استفاده از داده‌های آموزشی خود، آن‌ها می‌توانند یک «آداپتور LoRA» کوچک ایجاد کنند که با مدل Gemini Nano ترکیب شده و یک مدل زبان دقیق‌تر و شخصی‌سازی‌شده برای اپلیکیشن آن‌ها به وجود آورد.

https://android-developers.googleblog.com/2023/12/a-new-foundation-for-ai-on-android.html

11. ژن‌های پنهان:

https://blog.research.google/2023/05/building-better-pangenomes-to-improve.html

پانگنوم مجموعه‌ی کاملی از تمام انواع ژنتیکی یک گونه است که علاوه بر ژنوم مرجع، تنوع ژنتیکی موجود در جمعیت آن گونه را نیز دربرمی‌گیرد. دانشمندان با استفاده از هوش مصنوعی می‌توانند پانگنوم‌ها را بسازند و به درک بهتر تنوع ژنتیکی در گونه‌های مختلف برسند. با این حال، چالش‌هایی هم وجود دارد:

هزینه و زمان: توالی‌یابی DNA می‌تواند گران و زمان‌بر باشد.

پیچیدگی مونتاژ: ترکیب توالی‌های DNA، مخصوصاً در گونه‌هایی با تنوع ژنتیکی بالا، می‌تواند چالش‌برانگیز باشد.

تفسیر داده‌ها: درک و تفسیر اطلاعات پانگنوم می‌تواند دشوار باشد.

12. ترافیک هوایی

محققان Google AI با استفاده از یادگیری تقویتی عمیق، روشی برای بهبود کنترل ترافیک هوایی ارائه کرده‌اند. مدل یادگیری تقویتی از یک شبکه عصبی عمیق برای یادگیری «تابع ارزش» استفاده می‌کند که به آن می‌گوید هر وضعیت خاصی چقدر مطلوب است. سپس، با استفاده از الگوریتم «جستجوی درخت مونت کارلو» (MCTS)، بهترین اقدام را در هر وضعیت پیدا می‌کند. این روش با شبیه‌سازی تصادفی اقدامات مختلف و انتخاب بهترین گزینه، با وجود محاسبات سنگین، به بهبود عملکرد کنترل ترافیک هوایی کمک می‌کند.