بسم الله الرحمن الرحیم

محمد عرفان زارع زردینی

تمرین سری پنجم درس یادگیری عمیق

۹۸۴۱۱۴۳۲

مدرس درس: استاد داوود آبادی

زمستان 1402

(1

(آ) مسئله گربه بودن یا نبودن

یک شبکه عصبی کانولوشنال، وابسته بر لایه هایی است که تصاویر را بویله کانولوشن پردازش نموده و ویژگی ها را به صورت سلسله مراتبی استخراج می شود. این الگوها،لبه ها و اشکال درون تصویر را برای یادگیری بازنمایی ها تشخیص دهید. در مسئله گربه بودن یا نبودن، یک cnn می تواند ویژگی های خاصی همچون گوش، سبیل، بافت پوست و ساختار کلی بدن را آموزش بیند که نشان دهنده یک گربه است.

یک شبکه مبتنی بر توجه، وقتی برای این ویژگی طراحی شود، می تواند بر روی مناطق خاصی از تصویر تمرکز کند که بیشتر نشان دهنده و توصیف کننده ویژگی های گربه است. می تواند به صورت داینامیک توجه را به مناطقی با بافت های قابل توجه مانند موها، سبیل و قسمت های خاص بدن اختصاص دهد. این مکانیسم توجه به شبکه اجازه می دهد تا به طور انتخابی مناطقی را که برای شناسایی حضور یک گربه مورد نیاز هست پردازش کند و آن را در تشخیص بافت ها یا ویژگی های ظریف قوی تر نماید.

(ب) مسئله انسان بودن یا نبودن

یک CNN ویژگی هایی مانند نسبت اعضای صورت، چشم ها، بینی و دهان و الگوهای مشخص شده عناصر صورت را بیاموزد. البته، اگر تغییرات در نقاشی ها به طور قابل توجهی ساختار مرسوم چهره را نقض و از بین ببرد، CNN ها ممکن است مشکل داشته باشند. زیرا آنها در اصل از الگوها یاد می گیرند و ممکن است تغییرات شدید را نتوانند به خوبی مدیریت نمایند.

یک مدل مبتنی بر توجه برای این کار مناسب است زیرا می تواند به صورت داینامیک، به ویژگی های صورت مربوطه توجه کند. به عنوان مثال، می تواند بیشتر بر روی آرایش چشم ها، بینی و دهان تمرکز کند و اهمیت بیشتری به این مناطق بدهد. در مورد طراحی با عناصر صورت جابجا شده، یک مدل مبتنی بر توجه ممکن است با اختصاص سطوح مختلف توجه به قسمت‌های

مختلف تصویر، و تشخیص ناهنجاری‌های موجود و تصاویر نامرتبط در محل قرارگیری آنها، در تشخیص این اختلافات سازگارتر باشد.

(2

منبع:

https://chistio.ir/precision-recall-f/

الف)

مفاهیم اشاره شده مربوط به ماتریس سردرگمی اند که در یادگیری ماشین و آمار برای ارزیابی عملکرد مدل های طبقه بندی استفاده می شوند. توضیحات و عناوین آن به شرح زیر است:

:True Positive (TP)

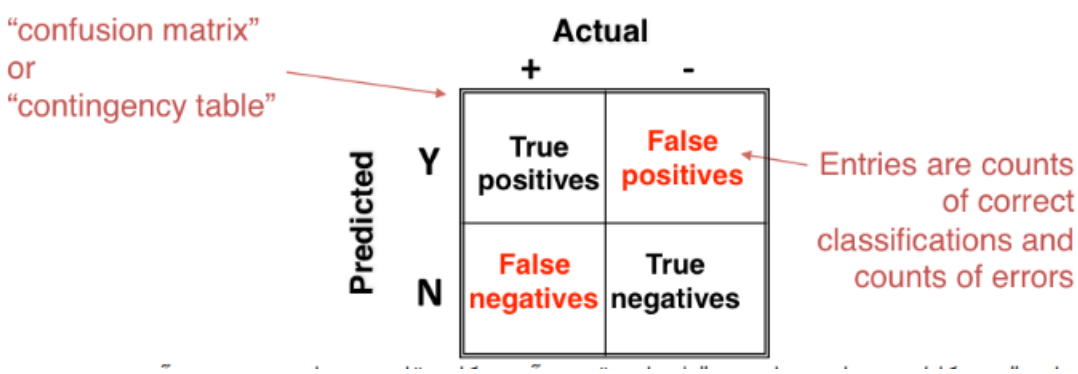
مواردی هست که مدل به درستی کلاس مثبت را پیش بینی می کند(یعنی نمونه های صحیح که به درستی دسته بندی شده اند). مثال در بخش ب، به درستی افراد درگیر در هک اسنپ فود به عنوان مجرم شناخته می شوند.

:True Negative (TN)

یک نمونه منفی به عنوان منفی دسته بندی شده است.(یعنی نمونه های اشتباه که به درستی به صورت اشتباه دسته بندی شده اند.) در اینجا شناسایی صحیح افرادی که در هک اسنپ فود دخیل نیستند به عنوان بی گناه می توان مثال باشد.

:False Positive (FP)

نمونه منفی به صورت نادرست به عنوان مثبت دسته بندی شده است.(نمونه اشتباهی که به اشتباه به عنوان صحیح دسته بندی شده است.) در اینجا شناسایی اشتباه افراد بی گناه به عنوان هکر است.



:False Negative (FN)

نمونه های مثبتی که به اشتباه ، منفی دسته بندی شده اند.(نمونه های صحیحی که به اشتباه به صورت اشتباه دسته بندی شده اند). در اینجا ناتوانی در شناسایی هکرهای واقعی است.

ب)

برای برقراری تعادل میان دقت و حفاظت و مصونیت افراد بی گناه ، معیار های اررزیابی زیر مهم اند:

Precision and Recall: Precision دقت پیش بینی های مثبت را می سنجد . در اینجا نسبت هکرهایی که به درستی شناسایی شده اند به همه هکرهای شناسایی شده می باشد. فرمول آن به شرح روبرو است.

$$\text{Precision (صحت)} = \frac{TP}{TP + FP}$$

Recall تعداد هکر های واقعی را که به درستی شناسایی شده اند را محاسبه می کند و فرمول آن به شرح زیر است:

$$\text{Recall} = \frac{TP}{TP + FN}$$

(پوشش)

متعادل نمودن این معیار بسیار مهم است زیرا دقت بالا ممکن است که سبب از دست رفتن برخی هکر ها شود(low recall) و بالعکس.

False Positive Rate: این معیار به درک میزان افراد بی گناهی که به اشتباه به عنوان هکر شناسایی شده اند کمک می کند. به حداقل رساندن این نرخ برای جلوگیری از دخالت افراد بی گناه بسیار مهم است.

F1 Score : این معیار رابط و وابسته precision ، recall است و زمانی کاربرد دارد که میان precision، recall بخواهیم تعادل برقرار کنیم. فرمول آن نیز به شرح زیر است:

$$\text{F1-score} = 2\frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = \frac{2\text{tp}}{2\text{tp} + \text{fp} + \text{fn}}.$$

دقت: یک معیار اساسی است، فرمول آن (پیش‌بینی‌های کل / پیش‌بینی صحیح)، و نباید تنها عامل باشد، زیرا ممکن است عملکرد واقعی را منعکس نکند، به خصوص در مجموعه داده‌های نامتعادل که تعداد هکرها ممکن است به میزان قابل توجهی پایین تر از کاربران بی گناه باشد.

بدین سان با تعادل میان معیارها ، ساختاری شکل می گیرد که سبب می شود هکر ها به شکل دقیق تری شناسایی شوند و در عین حال اتهامات نادرست علیه افراد بی گناه را به حداقل می رساند.

(3

الف)تخمین چرخش شامل پیش‌بینی جهت یا چرخش یک شی در یک تصویر است. می تواند برای کارهای طبقه بندی به روش های مختلف مفید باشد:

1- افزایش داده ها(Data Augmentation): با اعمال چرخش های تصادفی به داده های آموزشی، می توانید نمونه های آموزشی اضافی ایجاد کنید. این تکنیک تقویت به مدل کمک می‌کند تا با قرار دادن تصویر در جهت‌های مختلف اشیا، تعمیم بهتری پیدا کند و در بررسی های مختلف قوی‌تر شود.

2- تعمیم بهبودیافته(Improved Generalization): آموزش یک مدل برای پیش بینی چرخش یک شی در یک تصویر، آن را مجبور می کند تا ویژگی های انتزاعی بیشتری را بیاموزد. این می‌تواند به مدل در گرفتن بازنمایی‌های ثابت‌تر کمک کند و منجر به تعمیم بهتر به داده‌های دیده نشده شود.

3- Regularization: معرفی تخمین چرخش به عنوان یک کار کمکی در طول تمرین می تواند به عنوان یک تکنیک منظم سازی عمل کند. می‌تواند با تشویق مدل به یادگیری ویژگی‌های قوی‌تر و کلی‌تر از overfitting جلوگیری کند.

ب)

بردارهای one-hot ، راهی برای نمایش داده های دسته بندی به صورت عددی هستند. در این بردار، همه عناصر صفر هستند به جز یکی که نشان دهنده دسته است. به عنوان مثال، در یک سناریوی طبقه بندی دودویی، بردار [1، 0] ممکن است یک کلاس و [0، 1] کلاس دیگری را نشان دهد. مشکل بردارهای hot-one در ابعاد بالای آنها است، به ویژه در سناریوهایی با تعداد زیادی دسته آشکار می شود. به عنوان مثال، در یک کار طبقه بندی با 1000 کلاس، استفاده از رمزگذاری one-hot بردارهایی به طول 1000 ایجاد می کند. این ابعاد بالا می تواند منجر به ناکارآمدی محاسباتی، افزایش استفاده از حافظه و مشکلات در آموزش کارآمد شبکه های عصبی شود.

ج)

این روش یک الگوریتم self-supervised است که برای word embedding کاربرد دارد. این روش بدون داده های لیبل شده انسانی عمل می نماید و در عوض از مقادیر زیادی متن بدون

برچسب می آموزد. بوسیله مدل هایی همچون Continuous Bag of Words (CBOW)، Skip-gram ، word2vec یاد می گیرد که کلمات از دست رفته را باتوجه به ساختار آن پیش بینی کند یا کلمات متنی را از یک کلمه مرکزی در یک بخش متنی پیش‌بینی کند. این رویکرد، بازنمایی معنی‌داری از کلمات را با به تصویر کشیدن روابط آنها بر اساس نحوه حضور آنها در جملات استخراج می‌کند. توانایی Word2Vec برای تولید این نمایش‌ها بدون تکیه بر برچسب‌های خارجی، با قاعده اصلی یادگیری خود نظارتی ، با استفاده از ساختار ذاتی داده‌ها برای ایجاد بازنمایی‌های ارزشمند مطابقت دارد.

4)

الف)

پیرامون جستجوی معماری عصبی (NAS)، یادگیری تقویتی (RL) برای خودکار کردن فرآیند کشف معماری‌های شبکه بهینه استفاده می‌شود. عملکرد آن در این مورد به صورت زیر میتوان اشاره نمود:

1- تعامل agent و محیط: در NAS مبتنی بر RL، فرآیند جستجوی معماری شبکه به عنوان عاملی در تعامل با یک محیط پارامتربندی می شود. عامل مان یک فضای جستجو از معماری های ممکن را بررسی می کند و ساختارهای شبکه را تغییر می دهد یا ایجاد می نماید.
2- عملکرد agent و فیدبک محیطی: عامل حرکتی را انتخاب می‌کند که نشان‌دهنده تغییرات معماری هستند، مانند افزودن یا حذف لایه‌ها، تنظیم اتصالات یا اصلاح عملیات در شبکه عصبی.
3- سیگنال پاداش: پس از انجام یک حرکت، عامل فیدبکی را از محیط به صورت سیگنال پاداش دریافت می کند. این سیگنال پاداش عملکرد شبکه عصبی را با معماری اصلاح شده می سنجد. به عنوان مثال، پاداش ممکن است بر اساس دقت شبکه در یک مجموعه داده اعتبارسنجی یا سایر معیارهای عملکرد باشد.

4- Policy learning: عامل از تکنیک های یادگیری تقویتی، مانند policy gradients یا Q-learning ، برای به روز رسانی استراتژی خود برای انتخاب عملکرد ها استفاده می کند. هدف، به حداکثر رساندن پاداش تجمعی نسبت به تغییرات معماری متعدد است، و عامل را به سمت معماری هایی هدایت می کند که در وظیفه داده شده بهتر عمل می کنند.

5- Exploration و Exploitation : این الگوریتم ، میان دو مورد بیان شده برای حرکت موثر در فضای جست و جو وسیع معماری های ممکن ، تعادل برقرار می کند.

ب)

روش ذکر شده در قسمت قبل می تواند بهینه شدن اندازه تصویر ورودی و تعداد لایه ها برای مدل های تشخیص شی مورد استفاده است. حال به بررسی موارد گفته شده می پردازیم:

1- تصویر ورودی:

ارتباط رویکرد جست وجو: در nas میتوان اندازه های مختلف تصویر ورودی را برای مشخص نمودن اندازه بهینه برای کار تشخیص شی بررسیی کرد. اندازه های تصویر مختلف می تواند بر میدان های دریافت مدل و سطح جزئیات گرفته شده و نیازهای محاسباتی اثر بگذارد.

دلیل اهمیت داشتن: اندازه ورودی بزرگتر ممکن است که جزئیات پیچیده تری را ثبت نماید اما هرینه محاسباتی بیشتری دارد. در عوض، اندازه های کوچکتر ممکن است اطلاعات مهم را از دست بدهند. NAS می تواند تعادل بین جزئیات و کارایی را برای دقت تشخیص بهینه جستجو کند.

2- تعداد لایه ها:

ارتباط با رویکرد جست وجو: NAS می تواند تعداد ایده آل لایه ها را در معماری عصبی جستجو کند. عمق و پیچیدگی های مختلف می تواند بر توانایی مدل در استخراج ویژگی ها و تعمیم به خوبی به داده های دیده نشده، تأثیر بگذارد.

دلیل اهمیت داشتن آن: شبکه های عمیق تر می توانند نمایش های پیچیده ای را ثبت کنند اما ممکن است توانا در افزایش overfitting یا افزایش تقاضاهای محاسباتی باشند. شبکه های کم عمق ممکن است بهتر تعمیم دهند، اما ممکن است فاقد ظرفیت یادگیری ویژگی های پیچیده باشند. NAS می تواند عمق بهینه را برای متعادل کردن عملکرد و کارایی پیدا کند.

درواقع با استفاده از NAS، فرآیند جستجو شامل بررسی طیف وسیعی از احتمالات برای اندازه‌های ورودی و پیکربندی‌های معماری است. با استفاده از استراتژی‌های RL یا تکاملی، الگوریتم جستجو می‌تواند در این فضای احتمالی حرکت کند، عملکرد (به عنوان مثال، دقت، هزینه محاسباتی) را بررسی کرده و استراتژی exploration خود را به‌روزرسانی کند تا بر روی ترکیبی بهینه از اندازه تصویر ورودی و پارامترهای معماری عصبی همگرا شود. این به تطبیق مدل با نیازهای خاص در تشخیص شی کمک نموده و عملکرد را به حداکثر رسانده و در عین حال سربار محاسباتی را به حداقل می رساند .

(5

منبع:chatgpt(when training a standard GAN network, the value of the generating and critical loss function at the end of the first and 100th epoch are almost the same, why the quality of the images produced in the first and 100th epoch are not necessarily the same?)

در آموزش این شبکه، شبکه مولد و تفکیک کننده به صورت تخاصمی آموزش داده می شوند. هدف مولد تولید تصاویر واقعی برای فریب دادن تفکیک کننده است، در حالی که تفکیک کننده سعی می کند بین تصاویر واقعی و جعلی تمایز قائل شود. توابع از دست دادن برای هر دو شبکه به هدایت یادگیری آنها در طول آموزش کمک می کند. حتی در صورتی که مقادیر تلفات در پایان دوره اول و 100م مشابه باشد، کیفیت تصاویر تولید شده می تواند به دلایل مختلفی همچون موارد زیر متفاوت باشند:

1-یادگیری پویا: تابع ضرر معیاری از عملکرد را بر اساس اهدافمان ارائه می دهند. البته در حالی که فرآیند یادگیری را هدایت می کنند، ممکن است پیچیدگی کامل پویایی یادگیری را درک

نکند. شبکه ها ممکن است بدون دستیابی به همان سطح پیچیدگی در تولید تصویر، به مقادیر تلفات مشابه همگرا شوند.

2- Stochasticity در آموزش: آموزش GAN شامل درجه ای از تصادفی بودن، هم در انتخاب دسته ای از داده ها و هم در به روز رسانی وزن شبکه است. این تصادفی می تواند منجر به مسیرهای مختلف همگرایی، حتی با مقادیر تلفات مشابه شود. کیفیت تصاویر تولید شده را می توان تحت تأثیر تغییرات تصادفی مختلفی که در طول آموزش با آنها مواجه می شود، در نظر گرفت.

3-Collapse or Oscillation : GAN ها می توانند از این موارد و مشکلات رنج ببرند، که در آن تولیدکننده فقط تنوع محدودی از تصاویر را تولید می کند و تنوع موجود در داده های آموزشی را نادیده می گیرد که می تواند در هر نقطه ای از آموزش اتفاق بیفتد و کیفیت تصویر را با وجود مقادیر تلفات به ظاهر پایدار تحت تاثیر قرار دهد. نوسانات بین حالت ها یا همگرایی به حداقل محلی نیز ممکن است بر کیفیت تصویر تأثیر بگذارد.

4- آموزش اولیه دربرابر اصلاحات بعدی:

در epochهای اولیه، شبکه ممکن است یاد بگیرد که الگوهای ساده یا ویژگی های اساسی تولید کند که در نتیجه تصاویر با کیفیت پایینی تولید می شود. با پیشرفت آموزش، مدل ممکن است جزئیات و پیچیدگی های پیچیده تری را در داده ها ثبت کند که منجر به بهبود کیفیت تصویر می شود.

5- ابر پارامترها و معماری مدلمان: تغییرات در هایپرپارامترها یا اصلاحات در معماری شبکه در طول آموزش (مانند افزودن لایه ها، تنظیم نرخ یادگیری) می تواند بر روند یادگیری تأثیر بگذارد و سبب تاثیر بر کیفیت تصویر شود.

با وجود اینکه مقدار ضرر ،نشانه مناسبی برای همگرایی هست، ممکن است به طور کامل، کیفیت یا تنوع تصویرها، نمایش داده نشوند. پس تفاوت در کیفیت تصاویر میان دوره اول و100م می تواند به دلیل مسیر آموختن شبکه، تصادفی بودن آموزش، یا محدودیت های ذاتی آموزش پویا GAN باشد.