هوش ماشین طبیعت

مقاله

https://doi.org/10.1038/s42256-023-00647-z

پیش آموزش ساختار القا کننده

متیو بی ای مک درموت [1,2] ، برندان یاپ [1] ، پیتر شولوویتس [1] و مارینکا زیتنیک [2,3,4]

تاریخ دریافت: 12 مرداد 1401

پذیرش: 02 اسفند 1401

منتشر شده آنلاین: 1 ژوئن 2023

بررسی به روزرسانی

پیش آموزش مدل زبان و روش های همه منظوره مشتق شده، تحقیقات یادگیری ماشین را تغییر داده است. با این حال، عدم قطعیت قابل توجهی در مورد اینکه چرا پیش از آموزش عملکرد وظایف پایین دستی را بهبود می بخشد، وجود دارد. این چالش هنگام استفاده از پیش آموزش مدل زبان در حوزه های خارج از زبان طبیعی تلفظ می شود. در اینجا ما این مشکل را با تجزیه و تحلیل چگونگی تحمیل ساختار رابطه ای در فضاهای نهان القایی در هر نمونه بررسی می کنیم - یعنی روش های پیش آموزش چه محدودیت هایی را بر فاصله یا هندسه بین جاسازی های از پیش آموزش دیده نمونه ها اعمال می کنند. بررسی جامع روش های پیش از آموزش نشان می دهد که این سوال با وجود تحلیل های نظری که اهمیت درک این شکل از ساختار القایی را نشان می دهد، باز است. بر اساس این بررسی، ما یک چارچوب پیش از آموزش را معرفی می کنیم که درک دقیق و جامعی از چگونگی القای ساختار رابطه ای را امکان پذیر می کند. ما یک تجزیه و تحلیل نظری از چارچوب از اصول اولیه ارائه می دهیم و ارتباطی بین سوگیری استقرایی رابطه ای برقرار می کنیم
عملکرد قبل از آموزش و تنظیم دقیق. مطالعات تجربی شامل سه روش داده و ده وظیفه تنظیم دقیق، تجزیه و تحلیل های نظری را تأیید می کند، طراحی روش های جدید پیش آموزش را اطلاع رسانی می کند و پیشرفت های مداوم را در مجموعه ای از روش ها ایجاد می کند.

به عنوان مثال، در NLP، ارزیابی احساسات یک بررسی کامل رستوران یک کار به ازای هر نمونه است، برخلاف شناسایی یک توکن موجودیت نامگذاری شده در یک جمله، که یک کار درون نمونه و هر نشانه است. یکی از جنبه های PT که چنین عملکرد نهایی FT را هدایت می کند، هندسه القایی فضای نهفته از پیش آموزش دیده و در هر نمونه XX است (که به طور رسمی در Meth-ods تعریف شده است). به عنوان مثال، به خوبی مستند شده است که جاسازی جملات تولید شده توسط مدل های زبانی از پیش آموزش دیده در NLP می تواند غیر روان و ناهمسانگرد باشد، که به عملکرد کار پایین دست آسیب می رساند[توجه داشته باشید که استفاده ما از اصطلاح مدل زبان به روش هایی اشاره دارد که برای تولید جاسازی ها یا فعال کردن FT از مدل های زبان از پیش آموزش دیده طراحی شده اند، نه به مدل های زبان خودهمبسته برای نسل). در حوزه های دیگر ، مانند روش های زیست پزشکی، که وظایف هر نمونه حتی بیشتر از وظایف درون نمونه ای در مقایسه با NLP است، اهمیت این هندسه فقط افزایش می یابد. با وجود این اهمیت، تحقیقات در مورد مکانیسم های القای محدودیت های ساختاری صریح و عمیق در XX محدود است. به عنوان مثال، بسیاری از روش ها هندسه را نادیده می گیرند

پارادایم یادگیری پیش از آموزش (PT)/تنظیم دقیق (FT) (همچنین به عنوان یادگیری انتقال شناخته می شود) تأثیر فوق العاده ای بر پردازش زبان طبیعی (NLP) و حوزه های مرتبط [1-3] داشته است. روش های PT/FT مدل هایی را تولید کرده اند که قادر به ارائه پاسخ های متن آزاد به سؤالات زبان طبیعی[4]، پیش بینی خواص پروتئین ها از توالی[های 5] و امکان پیش بینی سنتز واکنش از رشته های 6 سیستم ورودی خط ورودی مولکولی ساده شده مولکولی 6 (SMILES) هستند.

در NLP یا PT/FT مشتق شده از NLP، برای یک روش داده پیش از آموزش معین

XX ، یک مجموعه داده $X \in XXNPT$ با اندازه $NPT \in Z$ به ما داده می شود و یک رمزگذار $f\theta : XXX\ XX$ پارامتریزه شده توسط $\vec{\theta}$ *را از قبل آموزش می دهیم* ، که XX را به یک فضای نهفته XX ترسیم می کند. سپس این رمزگذار $f\theta$ برای استفاده در کارهای مختلف FT (که در طول PT مشخص نیستند) منتقل می شود. ما سیستم های PT/FT را از طریق عملکرد $f\theta$ در وظایف FT مذکور ارزیابی می کنیم.

در این مقاله، ما در درجه اول به اثربخشی PT/ FT برای کارهای پایین دستی که در سطح هر نمونه کار می کنند، می پردازیم. برای

1آزمایشگاه علوم کامپیوتر و هوش مصنوعی، موسسه فناوری ماساچوست، کمبریج، MA، ایالات متحده آمریکا. 2گروه انفورماتیک زیست پزشکی، دانشکده پزشکی هاروارد، بوستون، MA، ایالات متحده آمریکا. 3موسسه برود MIT و هاروارد، کمبریج، MA، ایالات متحده آمریکا. 4بنکار علوم داده هاروارد، کمبریج، MA، ایالات متحده آمریکا. الکترونیکی: marinka@hms.harvard.edu

این منطقه مورد توجه ویژه ای است زیرا روش ها در NLP 1[3,24] موفق بوده اند، تعداد زیادی از روش های مشتق شده را در روش های غیرزبانی و زیست پزشکی[25-28] برانگیخته اند و هنوز از نظر فنی به طور کامل درک نشده‌اند.

تعریف محدودیت های ساختاری صریح و عمیق

در مرکز فرضیه ما این ادعا وجود دارد که امروزه اکثر متئودهای PT مشتق شده از NLP محدودیت های صریح و عمیقی را بر هندسه فضای نهان (در هر نمونه) XX تحمیل نمی کنند. برای توجیه این ادعا، محدودیت های ساختاری صریح و عمیق را از طریق تعاریف زیر تعریف می کنیم.

تعریف 1 محدودیت های ساختاری صریح در مقابل ضمنی. یک *LPT* هدف PT *یک محدودیت ساختاری را تحمیل می کند که صریح (در مقابل ضمنی) است تا حدی که (همانطور که* f_θ *به بهینه نزدیک می شود) به ما اجازه می دهد تا مستقیما در مورد رابطه (به ویژه فاصله) بین هر دو نمونه* **zi** *و* **zj** *در فضای نهفته* XX *استدلال کنیم، جایی که زیرنویس های* i *و* j *صرفا برای تمایز بین این دو نمونه در* XX *استفاده می شوند.* **تعریف 2 محدودیت های ساختاری عمیق در مقابل محدودیت های ساختاری کم عمق.**

یک LPT هدف PT یک محدودیت ساختاری را تحمیل می کند که عمیق (در مقابل کم عمق) بر اساس میزان اطلاعات (به عنوان مثال، چند بعد) برای برآورده کردن کامل محدودیت مورد نیاز است.

به عنوان مثال، یک طبقه بندی از دست دادن PT را با برچسب های موجود در مجموعه در نظر بگیرید YY، با نمونه i دارای برچسب $y_i \in$ YY، و با استفاده از یک لایه لاجیت که نمایش القایی نمونه i را به یک امتیاز پیش بینی **شده ترسیم می کند:** $\mathbf{z}_i \mapsto \tilde{y}_i$. این روش یک محدودیت ساختاری صریح ایجاد می کند زیرا ، نزدیک به بهینه بودن ، می توانیم استنباط کنیم که فاصله نسبی (کسینوس) بین دو نمونه **zi** و **zj** کوچک است اگر و فقط اگر $y_i = y_j$ باشد. با این حال، این محدودیت نیز کم عمق است زیرا برای برآورده کردن کامل این محدودیت، فقط باید هر کلاس $c \in$ YY را با یک موقعیت منحصر به فرد **کامپیوتر** $\ni$ XX جاسازی کنیم، سپس تمام sam- را فشرده کنیم.

zi Ples در نزدیکی نمونه اولیه کلاس خود $\mathbf{py}_I$. علاوه بر این، این محدودیت مبتنی بر فاصله را می توان در یک فضای بسیار کم بعد XX انجام داد (برای به عنوان مثال، ما می توانیم هر **کامپیوتر** را به طور یکنواخت در مورد یک دایره واحد دو بعدی توزیع کنیم، سپس تمام **Zi** را فشرده کنیم تا در حداقل فاصله کسینوس از نمونه های اولیه کلاس خود ظاهر شوند)، نشان می دهد که این محدودیت بسیار کم عمق است.

در مقابل، یک روش تضادی را در نظر بگیرید که ادعا می کند $\mathbf{z} = f(\mathbf{x})$ باید نزدیک به $\mathbf{z} = f(\mathbf{x})$ باشد، جایی که $\mathbf{x}$ یک نسخه آشفته است

مشتق شده از روش های PT NLP علاقه مند هستیم. این از XX با تحمیل هیچ از دست دادن PT بر روی تعبیه های کل نمونه[3,8,9]. روش های دیگر یا فقط محدودیت های کم عمق را تحمیل می کنند، مانند از طریق طبقه بندی کمکی هدف PT[1,10,11]، یا محدودیت های ساختاری عمیق تر، اما به شیوه ای ضمنی، مانند از طریق افزایش داده[های مبتنی بر افزایش داده 12-17] یا تلفات تضادی مبتنی بر نویز[18,19]. در حالی که چنین روش هایی می توانند قدرتمند باشند و در بسیاری از زمینه ها موفق بوده اند، ما استدلال می کنیم که فقدان یک چارچوب روشن برای طراحی روش های PT که محدودیت های ساختاری را بر XX تحمیل می کند که به طور همزمان صریح (مشابه تلفات طبقه بندی نظارت شده) و عمیق (مشابه تلفات تضادی مبتنی بر سر و صدا و مبتنی بر تقویت) هستند، یک ضعف اساسی است.

بر اساس این مشاهده، ما چارچوبی را توسعه می دهیم که بر اساس آن هدف PT به دو جزء تقسیم می شود: اول، یک هدف جانهی مدل زبانی یا نویز زدایی که از روابط درون نمونه استفاده می کند، و دوم، یک اصطلاح ضرر که برای منظم کردن هندسه فضای نهان هر نمونه XX هدایت می شود تا الگوهای اتصال یک نمودار مشخص شده توسط کاربر را منعکس کند. با تکیه بر نمودارها برای گرفتن ساختاری که می خواهیم در XX القا کنیم، این چارچوب به ما امکان می دهد روش های PT را مشخص کنیم که ساختار عمیق را به شیوه ای صریح القا می کنند و دقیقا شکاف مشخص شده در بالا را پر می کنند. علاوه بر این ، این پارادایم می تواند روابط متنوعی را به تصویر بکشد ، مانند روابط با انگیزه دانش خارجی (به عنوان مثال ، ref. 20)، محدودیت های خود نظارتی (به عنوان مثال، refs. 21,22) یا فاصله بین نمونه ها در یک حالت جایگزین (به عنوان مثال، ref. 23). علاوه بر این، این چارچوب PT به طور همزمان خاص است تا به ما امکان می دهد تضمین های نظری در مورد چگونگی تأثیر نمودارهای مختلف PT بر عملکرد FT ارائه دهیم، به اندازه کافی کلی که انواع روش های PT را در بر بگیرد و به اندازه کافی بیانگر برای ایجاد انگیزه در روش های جدید PT است که قبلا مورد مطالعه قرار نگرفته اند. علاوه بر تجزیه و تحلیل نظری، ما به صورت تجربی نشان می دهیم که تعریف روش های جدید با توجه به چارچوب ما، با استفاده از اشکال صریح ساختار دنیای واقعی، مزایای قابل توجهی نسبت به خطوط پایه PT رقابتی در سه روش و ده وظیفه FT به همراه دارد.

کار ما تحقیقات PT/FT را از طریق سه مشارکت پیش می برد. ابتدا از طریق یک بررسی جامع و تفسیر دقیق، نشان می دهیم که روش های PT موجود محدودیت های ساختاری ایجاد نمی کنند بیش از XX که به طور همزمان عمیق و صریح هستند. دوم، ما ایجاد می کنیم چارچوبی برای توصیف روش های PT، که مکانیزمی را برای طراحی روش های PT فراهم می کند که به صراحت محدودیت های ساختاری عمیق را در XX توسط یک نمودار PT مشخص شده توسط کاربر ایجاد می کند. ما بیشتر از این چارچوب با نتایج نظری پشتیبانی می کنیم که نحوه ارتباط ساختار نمودار با عملکرد وظیفه FT را تعیین می کند. مهمتر از همه، این رسمیت بخشیدن در پارادایم جدید PT ما بینشی را در مورد اینکه چه زمانی PT به تنهایی نسبت به یادگیری تحت نظارت ارزش افزوده می کند یا نمی کند، ارائه می دهد. سوم، ما نشان می دهیم که روش های PT القا کننده ساختار از طریق چارچوب ما در سطح یا بالاتر از روش های PT موجود در سه روش داده و ده وظیفه FT عمل می کنند.

نتیجه

فرمول بندی عمومی مسئله PT

با توجه به مجموعه داده $\mathbf{XPT} \in XX^{NPT}$، یک روش PT با هدف یادگیری یک رمزگذار XXX XX : $f\theta$ به گونه ای که $f\theta$ را می توان به وظایف FT که در زمان PT ناشناخته هستند منتقل کرد. در حالی که ما می توانیم از اطلاعات اضافی در زمان PT برای اطلاع رسانی از آموزش $f\theta$ استفاده کنیم (به عنوان مثال، برچسب های مخصوص YPT PT)، رمزگذار fθ *باید فقط نمونه هایی از* XX را به عنوان ورودی بگیرد تا بتوان از آن برای FT استفاده کرد. . به عنوان مثال، در مدل نمایش رمزگذار دو طرفه از ترانسفورماتورها ،(BERT) XX شامل نمونه های متن آزاد، $f\theta$ یک مدل ترانسفورماتور و LPT شامل هر دو مدل سازی زبان پوشانده شده از دست دادن هر نشانه و پیش بینی جمله بعدی (NSP) در هر نمونه[1] است.

تعریف ما از PT کاربردهای ثانویه هدف PT را نادیده می گیرد؛ به عنوان مثال، مدل های زبان خودرگرسیون (برای امتحان، ترانسفورماتور از پیش آموزش دیده مولد ([3] ref. 3-(GPT)) اغلب برای استفاده مولد آنها به طور مستقیم و نه به طور معمول برای به دست آوردن جاسازی ها یا در یادگیری انتقال استفاده می شود. بنابراین، ما به روش های PT

https://doi.org/10.1038/s42256-023-00647-z

از $\vec{x}_i$ تحت برخی از روش های سر و صدا یا تقویت $\mathbf{x}_i \mapsto \tilde{\mathbf{x}}_i$ ، اما به طور همزمان دور از سایر نمونه **های zj**. در حالی که این روش فضای نهان را محدود می کند تا با توجه به فرآیند نویز صاف باشد، فقط یک محدودیت ضمنی در XX ارائه می دهد، زیرا به طور کلی نمی توان استنباط کرد که چگونه فاصله بین نمونه های متمایز **zi** و **zj** محدود شده است. با این حال، محدودیتی عمیق تر از طبقه بندی اجباری تحمیل می کند، زیرا ارتباطات ضمنی بین نمونه های ناشی از روش نویز منعکس کننده روابطی است که لزوما نمی توانند در یک فضای کم بعد (بسته به اندازه و چگالی مجموعه داده) ثبت شوند.

محدودیت های متد PT موجود

برای نشان دادن اینکه روش های موجود به طور گسترده ابزاری برای تحمیل محدودیت های ساختاری که به طور همزمان عمیق و صریح هستند، ارائه نمی دهند، ما بیش از 90 روش PT موجود را بر اساس اینکه چگونه توابع هدف آنها XX را محدود می کند، بررسی می کنیم (داده های توسعه یافته شکل 1 و اطلاعات تکمیلی تکمیل). برای جزئیات کامل در مورد یافته های مرور ما، به روش ها مراجعه کنید. در تمام روش های بررسی شده، متوجه می شویم که محدودیت های ساختاری عمیق و صریح به ندرت مورد استفاده قرار می گیرند. در عوض، اکثر روش ها یا (1) هیچ هدف PT برای هر نمونه را تحمیل نمی کنند (به عنوان مثال، مدل های تولید متن، که اغلب برای جاسازی ها استفاده نمی شوند، بلکه برای برنامه های کاربردی محرک یا مولد[3,8,9,31] استفاده می شوند).

(2) از اهداف PT صریح، اما کم عمق و تحت نظارت استفاده کنید (به عنوان مثال، هدف NSP BERT، هدف محمول ترتیب جمله A Lite BERT (SOP) (ALBERT) یا اهداف مختلف چند وظیفه[ای 1,10,11])، یا (3) از اهداف PT تضادی ضمنی، اما عمیق، بدون نظارت یا خود نظارت استفاده کنید (به عنوان مثال، تعبیه جمله متضاد losses[12,13,18,19,32] یا سایر رویکردهای مبتنی بر سر و صدا یا مبتنی بر تقویت[14-17]).

https://doi.org/10.1038/s42256-023-00647-z

LM یک هدف سنتی و درون نمونه ای است (به عنوان مثال، یک مدل زبانی)، و LSI یک هدف جدید و القا کننده ساختار است که برای منظم کردن هندسه فضای نهان در هر نمونه توسط روابط طراحی شده است

الگوریتم اتصال نزدیکترین همسایه تحت برخی از عملکرد از راه دور در XX GPT را بازیابی می کند (محدودیت رسمی در روش ها است). توجه داشته باشید که این محدودیت بین چارچوب ما و ثروت تحقیقات موجود متمرکز بر یادگیری نمایش نمودار [36-41] ارتباط برقرار می کند. این تکنیک ها در واقع بینش های ارزشمندی را در مورد چگونگی نمونه برداری از مینی بچ ها بر روی داده های ساختار یافته با نمودار و ابداع تلفات برای جاسازی نمودار ارائه می دهند؛ با این حال، بسیاری از روش ها برای مدل سازی داده های ساختار یافته نمودار، از جمله جاسازی گراف های عمیق نسبت داده شده و شبکه های عصبی کانولوشن گراف، نباید به عنوان جایگزینی برای تکنیک های ما در اینجا دیده شوند زیرا معمولا با زمینه هایی که در آن نمودار در استنتاج شناخته نشده است، سازگار نیستندزمان، و بنابراین نمی توان از آنها در تنظیمات PT ما استفاده کرد که در آن *fθ* باید فقط ورودی های XX را مستقیما دریافت کند.

از آنجایی که اصطلاح ضرر اضافه شده LSI به صراحت برای القای ساختار *GPT* در XX طراحی شده است، ما روش ها (به ویژه روش هایی که از محدودیت های ساختاری عمیق و صریح استفاده می کنند) را تحت روش های پیش آموزش ساختار القا کننده چارچوب (SIPT) آموزش می دهیم. بسیاری از رویکردهای PT موجود را می توان به عنوان روش های SIPT دوباره تحقق بخشید، از جمله اهداف PT مبتنی بر طبقه بندی مانند NSP یا SOP، روش های کنتراستیو، یا روش های تراز نمودار موجود (روش ها). اگرچه SIPT به گونه ای طراحی شده است که القای محدودیت های ساختاری عمیق و صریح را آسان تر کند، اما به اندازه کافی انعطاف پذیر است تا محدودیت های ساختاری ضمنی یا کم عمق را به دست آورد.

تحلیل های نظری

در چارچوب ما، می توان ساختار GPT گراف PT را به عملکرد نهایی کار FT مرتبط کرد. به طور خاص، همانطور که یک جاسازی *f* SIPT روی گراف *GPT* به بهینه سازی تحت LSI از دست دادن نزدیک می شود، یک فضای جاسازی ایجاد می کند به طوری که عملکرد نزدیکترین همسایه برای هر کار پایین دستی با عملکردی که می تواند از طریق الگوریتم نزدیکترین همسایه بر روی نمودار GPT به دست آید، محدود تر است . این واقعیت به طور مستقیم هندسه گراف *GPT* را با عملکرد نهایی FT یک جاسازی *f* SIPT *مرتبط* می کند. علاوه بر این، مزیت استفاده از یک محدودیت صریح را به جای یک محدودیت ضمنی نشان می دهد؛ با کنترل ساختار *GPT*، کاربران می توانند به طور مستقیم انتخاب کنند که سوگیری های استقرایی مختلف را به فرآیند PT اضافه کنند به گونه ای که تأثیر قابل اثباتی بر مناسب بودن نهایی برای وظایف پایین دستی FT داشته باشد.

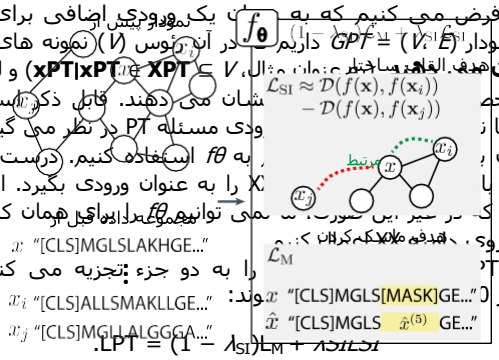
قضیه 1. *بگذارید* **XPT** *یک مجموعه داده PT باشد، بگذارید GPT یک نمودار PT باشد و بگذارید* f_θ *یک رمزگذار از پیش آموزش دیده تحت یک هدف PT مجاز در چارچوب ما باشد که مقدار* LSI *را بیش از I* درک نمی کند.سپس، تحت جاسازی f، دقت نزدیکترین همسایه برای یک وظیفه y FT همگرا می شود زیرا اندازه مجموعه داده ها حداقل به سازگاری محلی* (تعریف تکمیلی 3) *y over* G_{PT} افزایش می یابد.

ما دو نتیجه از قضیه 1 را ایجاد می کنیم که اهمیت انتخاب نمودارهای *GPT* را نشان می دهد که محدودیت های ساختاری عمیقی را تحمیل می کند.

اول، فرض می کنیم که به [illegible] یک ورودی اضافی برای مسئله PT، یک نمودار (V, E) *GPT* داریم که در آن رئوس (V) نمونه های PT را در **XPT نشان** [illegible] **xPT|xPT** $\in$ **XPT** $\subseteq V$، و لبه ها (E) روابط مشخص [illegible] نشان می دهند. قابل ذکر است که در حالی که ما [illegible] ورودی مسئله PT در نظر می گیریم، نمی توانیم از آن [illegible] *fθ* استفاده کنیم. درست مانند PT سنتی، *fθ* [illegible] XX را به عنوان ورودی بگیرد. این به این دلیل است که [illegible] نمی توانیم *fθ* را برای همان کلاس کلی وظایف FT روی [illegible]

دوم، [illegible] را به دو جزء تجزیه می کنیم که با هایپرپارامتر [illegible] اند:

$$\mathrm{LPT} = (1 - \lambda_{\mathrm{SI}})\mathcal{L}_{\mathrm{M}} + \lambda_{\mathrm{SI}}\mathcal{L}_{\mathrm{SI}}.$$



انجیر. 1 | چارچوب PT ما. ما فرمول PT را با در نظر گرفتن نمودار *GPT* PT به عنوان ورودی کمکی دوباره ریخته گری می کنیم.*GPT* برای تعریف یک LSI هدف القا کننده ساختار استفاده می شود، که یک رمزگذار *fθ* PT *را تحت فشار قرار می دهد* تا نمونه ها را جاسازی کند به گونه ای که نمونه ها در فضای نهفته نزدیک باشند، اگر و تنها در صورتی که در *GPT مرتبط باشند*.

(لبه ها) در جی پی تی. در چارچوب ما، LSI فقط برای جی پی تی، *fθ* و XX اگر اجازه می دهد برخی از بهینه سازی پایدار در آن نقطه شعاع

در تمام روش های بررسی شده، متوجه می شویم که تنها چهار روش محدودیت های صریح و عمیق را به طور همزمان تحمیل می کنند: تعبیه دانش و بازنمایی زبان انگلیسی از پیش آموزش دیده (KEPLER)[33]، GNN آگاه از دانش XLM-K[34]، (CK-GNN)[23] و WebFormer[35]. هر چهار را می توان به عنوان نوعی تراز نمودار در هر نمونه توصیف کرد، که در آن یک نمودار دانش PT *خارجی GPT* یا الگوریتم اتصال بر روی زیرمجموعه ای از نمونه های PT استفاده می شود و تعبیه های خروجی جفت نمونه **های** $f\theta(\mathbf{xi}) = \mathbf{zi}$ و $f\theta(\mathbf{xj}) = \mathbf{zj}$ محدود به انعکاس روابط خود در نمودار PT هستند. این شکل از محدودیت صریح است، زیرا گراف *GPT* حاوی روابط صریحی است که در فضای نهفته خروجی القا می شود، اما همچنین عمیق است، زیرا هندسه نمودار *GPT* می تواند به طور دلخواه پیچیده باشد.

با این حال، همه این روش ها محدودیت های عمده ای دارند. در KEPLER و XLM-K، جاسازی های هر نمونه فقط به مجموعه محدودی از نمونه های مربوط به توضیحات موجودیت از یک نمودار دانش محدود می شود. به این ترتیب، هیچ محدودیتی در نمونه های متن آزاد دامنه عمومی تنها در XX [33,34] وجود ندارد. در CK-GNN، اتصال گراف از یک گراف یک همسایه نزدیک به خوشه در فضای فاصله یک مدالیته جایگزین مشتق می شود، که ممکن است یک ساختار مرتبه بالاتر محدود را ارائه دهد. برخلاف رویکردهای NLP، این روش هیچ وظیفه PT 23 درون نمونه (به عنوان مثال، به ازای هر نشانه) ندارد. در نهایت، در WebFormer، نمودار مورد استفاده از ساختار صفحات وب زیربنایی زبان نشانه گذاری ابرمتن (HTML) استنباط می شود و روابط فقط در سطح هر نمونه برای روابط ساختاری محدود در HTML محدود می شوند. علاوه بر این، WebFormer یک مدل تخصصی به طور خاص برای پردازش محتوای وب (متن و عناصر HTML) است، بنابراین این رویکرد را نمی توان مستقیما به دامنه های دیگر تعمیم داد. علاوه بر این، این روش ها فقط زمینه های خاص مدل های خود را بررسی می کنند. آنها هیچ چارچوب کلی برای تحقق این محدودیت های عمیق و صریح در هر نمونه در زمینه های دیگر ارائه نمی دهند و هیچ نظریه ای را در مورد چگونگی ارتباط این محدودیت ها با عملکرد وظایف FT[23,33-35] بررسی نمی کنند.

به طور کلی، بررسی ما از روش های PT به صراحت نشان می دهد که روش های PT که قادر به ارائه محدودیت های ساختاری صریح و عمیق هستند، به طور قابل توجهی کمتر مورد بررسی قرار گرفته اند. در تمام روش هایی که بررسی کردیم، تنها چهار روش محدودیت های اهرم صریح و عمیق هستند که همگی محدودیت های قابل توجهی دارند و هیچ اجماعی در مورد چگونگی محدود کردن XX به طور صریح و عمیق وجود ندارد. این یافته ها به چارچوب ما انگیزه می دهد، که بینشی را برای تحقق محدودیت های ساختاری عمیق و صریح در مدل های PT در زمینه های مختلف ارائه می دهد و راهنمایی های نظری در مورد چگونگی ارتباط محدودیت های ساختاری با عملکرد FT ارائه می دهد. همانطور که در نتایج خود نشان می دهیم، القای محدودیت های عمیق و صریح از طریق چارچوب ما، مزایای قابل توجهی نسبت به متولوژی های PT موجود در سه حوزه زیست پزشکی متنوع ایجاد می کند.

PT القا کننده ساختار

چارچوب مسئله PT ما شامل دو تفاوت کوچک اما مهم با فرمول استاندارد است (شکل 1). 1).

https://doi.org/10.1038/s42256-023-00647-z

جدول **1** | خلاصه ای از مجموعه داده ها، وظایف و معیارهای ما

شبکه	چکیده ها	پروتئین	
نمودار ایگوگراف شبکه برهمکنش پروتئین-پروتئین	چکیده مقاله زیست پزشکی	توالی پروتئین	روش داده (xi یک ...)
Ref. ۲۶ عدد	نمودار آکادمیک مایکروسافت- بخش 21,22	درخت زندگی[20]	مجموعه داده PT
پروتئین مرکزی xi در همه به جز نه برچسب هستی شناسی ژن با پروتئین مرکزی xj موافق است.	مقاله شی به مقاله ایکس جی استناد می کند	xi با XJ تعامل دارد	$(xi, xj) \in GPT$
پوشش ویژگی[26]	SciBERT[53]	نوار[5]	خط پایه به ازای هر توکن
یادگیری چند وظیفه‌ای - بخش 26	BioLinkBERT[56]	به علاوه[52]	خط پایه به ازای هر نمونه
Ref. ۲۶ عدد	SciBERT[53]	نوار[5]	مجموعه داده FT

به عنوان مثال، برای دامنه پروتئین ها، مجموعه داده PT ما مجموعه ای از توالی های پروتئینی موجود در مجموعه داده درخت زندگی[20] است، پروتئین ها در GPT گراف PT ما مرتبط هستند اگر و تنها در صورتی که بر اساس نمودار درخت زندگی تعامل داشته باشند. علاوه بر این، ما وظایف FT را در معیار TAPE با خط پایه خام و هر توکن که به صورت عمومی در مدل TAPE[5] و پایه هر نمونه منتشر شده در مدل PLUS PT[52] در دسترس است، مقایسه می کنیم.

تضمین عملکرد FT، ناشی از میزان سازگاری برچسب وظیفه FT در کلاس های تحت هدف PT تحت نظارت است. در مقابل، اگر از یک محدودیت ساختاری عمیق استفاده شود، که در نتیجه 2 از طریق *GPT* که نزدیکترین نمودار همسایه بر روی یک منیفولد دلخواه N است، تحقق می یابد، یک مدل SIPT تضمین نظری برای عملکرد FT را مجاز می داند که با افزایش اندازه مجموعه داده PT برای هر کار FT که روی N صاف است، به وحدت نزدیک می شود.

این تحلیل نظری نشان می دهد که ما می توانیم به طور مستقیم ساختار القا شده در XX را به عملکرد FT پایین دست متصل کنیم. به این ترتیب، روش های جدید PT که از نمودارهای *GPT* با محدودیت های ساختاری عمیق تر استفاده می کنند، می توانند عملکرد را به طور قابل توجهی بهبود بخشند، همانطور که در مجموعه داده های دنیای واقعی در آزمایش های خود نشان خواهیم داد. اثبات کامل برای تمام نتایج تئورتیکال و آزمایش های نیمه سنتزی که یافته های نظری ما را در عمل تأیید می کند، در روش ها قرار دارد.

مجموعه داده ها و وظایف

ما سه روش داده را برای آزمایش های خود بررسی می کنیم: "پروتئین ها"، توالی های پروتئینی را حفظ می کنند. «چکیده ها»، حاوی چکیده های زیست پزشکی با متن آزاد ؛ و «شبکه ها»، حاوی زیرنگراف های شبکه های برهمکنش پروتئین-پروتئین (PPI).

در هر روش داده، ما از مجموعه داده های مختلف PT استفاده می کنیم و از انواع مختلف نمودارهای *GPT* PT *استفاده می کنیم*، معیارهای موجود برای عموم را برای وظایف FT آزمایش می کنیم و روش های SIPT خود را با خطوط پایه قانع کننده ای که هر دو روش هر نمونه و هر توکن را در بر می گیرد مقایسه می کنیم (جداول 1-3). جزئیات بیشتر در مورد این جنبه ها در روش ها آمده است.

LSI و روش های آموزشی

همانطور که در تعریف چارچوب ما بحث شد، یک روش SIPT با یک روش استاندارد PT با (1) انتخاب نمودار *GPT* (جدول 1) و (2) طراحی از دست دادن ساختار القا L. برای تعریف L در یک محدودیت ساختاری کم عمق استفاده می شود (به عنوان مثال، یک طبقه بندی تحت نظارت)، مدل معادل SIPT مرتبط فقط حداقل را مجاز می داند.

جدول **2** | میانگین (**±** انحراف معیار) کاهش نسبی خطا (تعریف شده به صورت ([خطای پایه] - [خطای مدل *GPT*])/[خطای پایه]) مدل های آموزش دیده تحت چارچوب ما در مقابل خطوط پایه منتشر شده به ازای هر نشانه یا هر نمونه

در مقابل هر نمونه		در مقابل PT به ازای هر توکن		وظیفه	دامنه
Δ	کاهش نسبی خطا	Δ	کاهش نسبی خطا		
↑	**8.4** درصد ± **2.4** عدد	↑	**7.0**٪ ± **1.۲** عدد	Rh	پروتئین
↑	**12.8**٪ ± **1.**۱ عدد	~	-0.8٪ ± 1.3	Fl	
~	2.8 ± ۲.۲٪	↑	**13.1**٪ ± **2.5** عدد	خیابان	
↑	**4.5**٪ ± **0.**۲ عدد	↑	**4.5**٪ ± **0.**۲ عدد	Ss	
	Na	↑	[a]**10.5**٪	Cp	
↑	**0.8** درصد ± **0.3** عدد	~	0.2 ± ۰.۳٪	Pf	چکیده ها
~	-5.5 ± ۱.۱٪	~	4.1 ± ۲.۴٪	Sc	
~	16.2 ± ۱۱.۶٪	↑	**17.7** درصد ± **6.5** عدد	Aa	
~	-10.1 ± ۳.۶٪	↑	**6.7** درصد ± **0.4** عدد	SRE	
↑	**5.1**٪ ± **2.7** عدد	~	5.2 ± ۷.۸٪		شبکه

اعداد بالاتر نشان می دهد که مدل های تحت چارچوب ما خطا را بیشتر کاهش می دهند و در نتیجه از خطوط پایه بهتر عمل می کنند. ستون Δ نشان می دهد که آیا مدل از نظر آماری بهبود معنی داری (↑ و پررنگ)، بدون تغییر معنی دار (~) یا کاهش آماری معنی دار (↓ و پررنگ) ارائه می دهد یا خیر. معنی داری آماری با استفاده از آزمون *t* در معنی داری ارزیابی می شود سطح $P < 0.1$. [a]تجزیه و تحلیل هر نمونه و برآورد واریانس برای CP به دلیل هزینه محاسباتی این کار غیرممکن بود. وظایف FT در جدول 3 توضیح داده شده است.

نتیجه 1. *بگذارید* $\mathbf{XPT} \in XXN$ *یک مجموعه داده PT با برچسب های مربوطه باشد* $G = (X, E)$. $\mathbf{y} \in Y^{N}$ *را به گونه ای تعریف کنید که* $(\mathbf{x}_i, \mathbf{x}_j) \in E$ *اگر و فقط اگر* $y^{(FT)}_{i} = y^{(FT)}_{j}$.

سپس، سازگاری محلی برای یک وظیفه FT **معین** $\mathbf{y}^{(FT)}$ *روی GPT (و بنابراین با قضیه 1، دقت نزدیکترین همسایه برای هر جاسازی کننده SIPT بهینه شده) با این احتمال که برچسب* $y^{(FT)}_{i}$ *با برچسب کلاس اکثریت برای وظیفه* $\mathbf{y}^{(FT)}$ *موافق باشد، محدود می شودبیش از دسته متشکل از همه گره ها با برچسب PT یکسان بله* y_i *به عنوان* x_i.

نتیجه 2. *بگذارید* **XPT** *یک مجموعه داده PT باشد که می تواند بر روی یک منیفولد معتبر N تحقق یابد. فرض کنید* **XPT** *با پشتیبانی کامل از N نمونه برداری شده است. اجازه دهید* $G_{PT}(\mathbf{XPT}, E)$ *یک نمودار r-نزدیکترین همسایه روی N باشد (به عنوان مثال* $(\mathbf{x}_i, \mathbf{x}_j) \in E$ *اگر و فقط اگر فاصله ژئودزیک بین دو نقطه روی باشد* $\mathrm{DDN}(\mathbf{x}_i, \mathbf{x}_j) < r$ *کمتر از r است:). بگذارید* $y^{(FT)}$ *یک کار طبقه بندی FT باشد که تقریبا در همه جا روی منیفولد صاف است.*

سپس، از آنجایی که اندازه مجموعه داده PT (و در نتیجه اندازه GPT) تمایل به ∞ دارد و r به صفر می رسد، سازگاری محلی $y^{(FT)}$ *بر روی GPT (و بنابراین توسط قضیه 1 دقت نزدیکترین همسایه یک جاسازی کننده SIPT) نیز به یک تمایل خواهد داشت.*

به طور غیررسمی، این نتیجه گیری ها ثابت می کنند که هنگامی که از

https://doi.org/10.1038/s42256-023-00647-z

آزمایشات، ما از ایده های یادگیری متریک حفظ ساختار[42-44] استفاده می کنیم. یادگیری متریک حفظ ساختار شکلی از یادگیری متریک است که در آن روابط مثبت توسط لبه های یک نمودار به جای یک برچسب تحت نظارت مشترک تعریف می شود. ما دو ضرر را تطبیق می دهیم، یک ضرر سنتی [45] و یک از دست دادن چند شباهت[46]، از یادگیری متریک نظارت شده تا زمینه مبتنی بر نمودار و حفظ ساختار اصطلاحات LSI در SIPT.

علاوه بر این تلفات، در حوزه های چکیده ها و پروتئین ها، ما از یک روش شروع گرم برای مقداردهی اولیه PT از مدل های زبانی موجود به جای شروع از صفر استفاده می کنیم. این باعث صرفه جویی در زمان محاسباتی قابل توجهی می شود و امکان یک مطالعه فرسایش قدرتمند را برای جداسازی بهبود عملکرد برای معرفی اصطلاح LSI ما فراهم می کند . دوم، ما مطالعات گسترده ای را در مورد تنظیم ابرپارامترها در این دو حوزه انجام می دهیم تا مقادیر مناسب برای λSI را شناسایی کنیم و آن یافته ها را با حوزه شبکه ها تطبیق دهیم. جزئیات بیشتر در مورد مجموعه آزمایشی، از جمله اظهارات رسمی از ضررهای کنتراستی و چند شباهت ما، در روش ها آمده است. توجه داشته باشید که همانطور که در برنامه های PT استاندارد است، برای هر الگوریتم PT و مدالیته داده، ما یک مدل واحد را از قبل آموزش می دهیم

https://doi.org/10.1038/s42256-023-00647-z

جدول **3** | وظایف **FT**

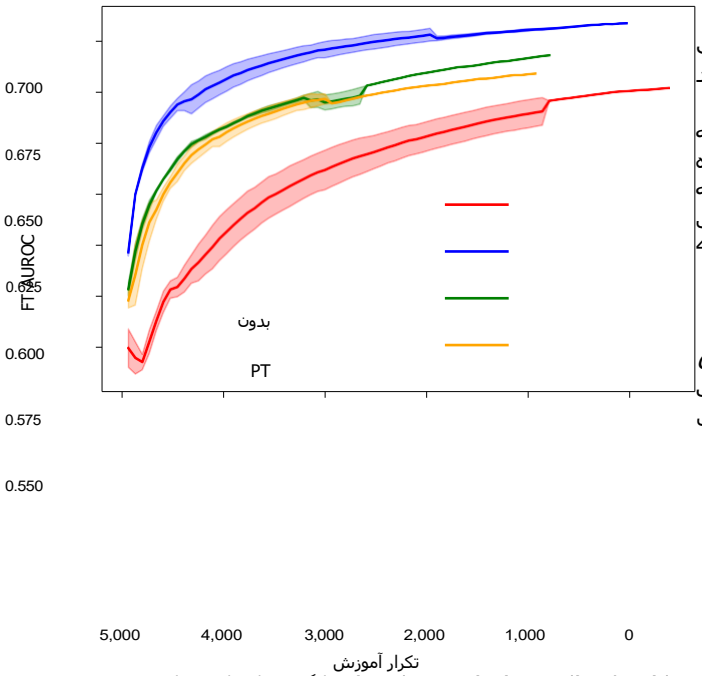
متریک	توضیحات	وظیفه FT		مجموعه داده FT
		نام	مخفف	
دقت	تکلیف طبقه بندی بر اساس توالی برای پیش بینی دسته چین های پروتئینی	همسانی از راه دور	Rh	نوار[5]
دقت	تکلیف طبقه بندی هر توکن برای پیش بینی خواص ساختاری اسید آمینه	ساختار ثانویه	Ss	
p/*سپرمن*	تکلیف رگرسیون پر-توالی برای پیش بینی پایداری	ثبات	خیابان	
p/*سپرمن*	تکلیف رگرسیون پر-توالی برای پیش بینی فلورسانس	فلورسانس	Fl	
دقت @ 5 / *L*	طبقه بندی درون دنباله ای برای پیش بینی اینکه کدام جفت ها اسیدهای آمینه در ترکیب سه بعدی پروتئین در تماس هستند	پیش بینی تماس	Cp	
ماکرو-F1	مسئله طبقه بندی هر جمله برای پیش بینی حوزه مطالعاتی یک مقاله از عنوان آن	زمینه مقاله	Pf	SciBERT[53]
ماکرو-F1	مسئله طبقه بندی هر جمله برای پیش بینی قصد استناد	SciCite	Sc	
ماکرو-F1	مسئله طبقه بندی هر جمله برای پیش بینی قصد استناد	ACL-ARC	Aa	
ماکرو-F1	استخراج رابطه هر جمله	استخراج رابطه علمی	SRE	
ماکرو-AUROC	طبقه بندی باینری چند برچسبی به 40 اصطلاح هستی شناسی ژن	شبکه‌ها[26]		

عملکرد فلورسانس (FL)، پایداری (ST) و میدان کاغذ (PF). برای جزئیات بیشتر در مورد این وظایف به جدول 3 و اطلاعات تکمیلی مراجعه کنید و به یاد بیاورید که متریک F1 میانگین هارمونیک دقت و یادآوری است.

شکل 2 نشان می دهد که چگونه عملکرد در تکرارهای FT برای مجموعه داده های شبکه ها تکامل می یابد تا مشخص شود که آیا بهبودهای مشاهده شده در مقادیر همگرای نهایی در طول آموزش وجود دارد یا خیر. می بینیم که روش های SIPT سریعتر به عملکرد بهتری نسبت به هر دو خط پایه همگرا می شوند . نتایج خام در تمام تنظیمات در جداول داده های توسعه یافته 3 و 4 ارائه شده است.

دستاوردهای عملکرد SIPT قوی است

دستاوردهای عملکرد SIPT در هر سه روش داده و انواع مختلف *GPT* ادامه دارد . این نشان می دهد که منظم کردن صریح هندسه فضای نهفته در هر نمونه ارزشی را در NLP، توالی های غیر زبانی و دامنه های غیر متوالی ارائه می دهد. بعلاوه



شکل 2 | عملکرد FT از طریق شبکه ها. میانگین ± انحراف معیار FT AUROC به عنوان تابعی از تکرار FT برای مجموعه داده شبکه ها. تفاوت در مقیاس واریانس ناشی از اجراهای مختلف است که باعث توقف زودهنگام در تکرارهای مختلف می شود. روش SIPT سریعتر همگرا می شود و عملکرد بهتری نسبت به درون نمونه (مدل سازی گره نقابدار) یا هر نمونه (طبقه بندی چند وظیفه) دارد. Mask-PT نشان دهنده انجام قبل از آموزش انتساب ماسک به تنهایی است، در حالی که SIPT ترکیبی از این دو رویکرد را از طریق چارچوب SIPT ما نشان می دهد.

کند. در برخی موارد، دستاوردهای عملکرد قابل توجه است، با بهبود تقریبا 17٪ (0.05 تغییر خام ماکرو-F1) در 6٪، ACL-ARC (AA) در استخراج رابطه SciERC (SRE) (0.01 تغییر مطلق ماکرو-F1) و 4٪ در همسانی از راه دور (RH؛ 2٪ تغییر دقت مطلق). مدل های SIPT عملکرد پیشرفته جدیدی را در AA و RH ایجاد می کنند و با پیشرفته ترین مدل ها مطابقت دارند

مجموعه داده PT، سپس آن یک مدل از پیش آموزش دیده را در هر وظیفه FT به طور مستقل تنظیم کنید؛ به عبارت دیگر، در هیچ تنظیمی نیازی به آموزش یک مدل جداگانه برای هر وظیفه FT نداریم.

SIPT با همه خطوط پایه مطابقت دارد یا بهتر عمل می کند

برای تجزیه و تحلیل آزمایشات خود، کاهش نسبی خطای مدل SIPT با بهترین عملکرد را در مقابل خطوط پایه هر توکن یا هر نمونه در تمام وظایف FT محاسبه می کنیم (جدول 2). در 10 مورد از 15 مورد، SIPT نسبت به روش های موجود بهبود می یابد. به هیچ وجه بدتر از هیچ یک از پایه ها عمل نمی

https://doi.org/10.1038/s42256-023-00647-z

استفاده از نمودارها، از جمله آنهایی که توسط دانش خارجی تعریف می شوند، توسط سیگنال های خود نظارتی در داده ها به طور مستقیم، و با روش های نزدیک ترین همسایه در فضاهای برچسب چند وظیفه ای، مفید است. علاوه بر این، این پیشرفت ها در مقایسه با رویکردهای مدل سازی زبان استاندارد و در برابر روش های موجود که اهداف PT را به ازای هر نمونه، از جمله اهداف طبقه بندی تک و چند وظیفه تحمیل می کنند، وجود دارد.

دستاوردها به LSI از دست دادن SIPT نسبت داده می شود

همانطور که در روش ها مشخص شده است، طراحی آزمایشی ما به ما اجازه می دهد تا تعیین کنیم که چه مقدار از دستاوردهای مشاهده شده در جدول 2 به دلیل مؤلفه از دست دادن SIPT است، برخلاف به عنوان مثال، آموزش مداوم، داده های PT جدید یا روش های انتخاب دسته ای مورد استفاده در روش ما، که به طور غیرمستقیم از دانش ذاتی GPT نیز استفاده می کند، که به طور غیرمستقیم از دانش ذاتی GPT نیز استفاده می کند.. جای تعجب نیست که برخی از دستاوردها به دلیل این عوامل دیگر مشاهده می شود و با توجه به این مطالعات فرسایش دستاوردهای عملکرد کاهش می یابد. با این حال، حتی هنگام مقایسه با حداکثر عملکرد پایه یا مطالعه فرسایش به طور کلی، نه جهت روابط مشاهده شده و نه اهمیت آماری مقایسه های مشاهده شده تغییر نمی کند. بنابراین، می توانیم به طور قطعی بیان کنیم که بهبودهای عملکردی مشاهده شده در اینجا به طور منحصر به فرد به اجزای القا کننده ساختار معرفی شده توسط چارچوب ما نسبت داده می شود. نتایج مطالعه فرسایش کامل را می توان در جداول داده های توسعه یافته 3 و 4 یافت.

بحث

با وجود گستردگی تحقیقات در مورد روش های PT، روش هایی برای اعمال محدودیت های ساختاری صریح و عمیق بر روی هر نمونه، PT

https://doi.org/10.1038/s42256-023-00647-z

وظیفه طبقه بندی برای پیش بینی دسته چین پروتئین (متریک: دقت)؛ ساختار ثانویه (SS)، یک وظیفه طبقه بندی به ازای هر توکن برای پیش بینی خواص ساختاری اسید آمینه (متریک: دقت)؛ پایداری (ST) و فلورسانس (FL)، هر توالی، وظایف رگرسیون برای پیش بینی پایداری و فلورسانس پروتئین، به ترتیب (متریک: Spearman's ρ)؛ و پیش بینی تماس (CP)، یک وظیفه طبقه بندی درون توالی برای پیش بینی اینکه کدام جفت اسیدهای آمینه در پروتئین در تماس هستند. ترکیب سه بعدی (متریک: دقت در $L/5$ که L طول پروتئین است). همه این وظایف از مجموعه داده های در دسترس عموم هستند که می توانند مستقیما در GitHub TAPE (https://github.com/songlab-cal/tape#data) به دست آیند، که هیچ مجوزی برای این مجموعه داده ها فهرست نمی کند، اگرچه GitHub کلی تحت مجوز BSD 3 بند "جدید" یا "تجدید نظر شده" منتشر شده است. RH مدلی را برای پیش بینی یک دسته چین پروتئین در سطح هر دنباله انجام می دهد. مجموعه داده این وظیفه شامل 718/736/12,312 پروتئین قطار/اعتبارسنجی/آزمایش است و در اصل از ref تهیه شده است. 47. SS یک مسئله طبقه بندی چند طبقه ای است که با استفاده از دقت ارزیابی می شود، که مدلی را برای پیش بینی خواص ساختاری هر اسید آمینه در پروتئین نهایی و تا شده انجام می دهد. مجموعه داده این وظیفه شامل 513/2,170/8,678 پروتئین قطار/اعتبارسنجی/آزمایش است و از ref تهیه شده است.

48. ST مدلی را برای پیش بینی پایداری پروتئین در پاسخ به شرایط محیطی انجام می دهد. مجموعه داده این وظیفه شامل 12,839/2,447/53,679 پروتئین قطار/اعتبارسنجی/آزمایش است که در اصل از ref تهیه شده است. 49. FL به مدلی نیاز دارد تا پیش بینی کند که یک پروتئین چقدر درخشش می کند. مجموعه داده این وظیفه شامل 27,217/5,362/21,446 پروتین قطار/اعتبارسنجی/آزمایش است و در اصل از ref تهیه شده است. 50. در نهایت، CP به مدلی نیاز دارد تا پیش بینی کند که آیا هر جفت اسید آمینه از یک پروتئین کمتر از 8 Å از هم فاصله دارند یا خیر. مجموعه داده این وظیفه از ProteinNet[51] تهیه شده است. در این آزمایشات، ما با مدل TAPE منتشر شده 5 مقایسه می کنیم، که از یک وظیفه مدل سازی زبان به تنهایی به عنوان نقطه مقایسه هر نشانه استفاده می کند، و مدل نمایش توالی پروتئین آموخته شده با استفاده از اطلاعات ساختاری (PLUS)[52]، که برای LM بهینه سازی می کند و طبقه بندی نظارت شده به طور مشترک برای مقایسه هر نمونه ما

نقطه.

مجموعه داده Abstracts و تسک های FT

ما از مجموعه داده ای از ~650,000 چکیده مقاله علمی متن آزاد از مجموعه داده گراف آکادمیک مایکروسافت (MAG)[21,22] استفاده می کنیم. داده های Abstracts PT (مجموعه داده MAG) دارای مجوز با انتساب Open Data Commons هستند

مجوز (ODC-By) نسخه 1.0.

دو چکیده در *GPT* برای این مجموعه داده مرتبط هستند اگر و فقط در صورتی که مقالات مربوطه آنها به یکدیگر استناد کنند. این یک نمودار خود نظارتی است.

در اینجا، ما از زیرمجموعه ای از وظایف FT مورد استفاده در مقاله SciBERT[53] استفاده می کنیم، از جمله فیلد کاغذی ACL- (PF)، SciCite (SC)، ARC (AA) و استخراج رابطه SciERC (SRE)، که همگی مسائل طبقه بندی هر جمله هستند (متریک: ماکرو-F1). مدل های وظایف PF برای پیش بینی حوزه مطالعاتی یک مقاله از عنوان آن، وظایف SC و AA هر دو برچسب "قصد" را برای استنادها پیش بینی می کنند و SRE یک کار استخراج رابطه است. تمام مجموعه داده های FT را می توان از SciBERT GitHub (https://github.com/allenai/scibert)، که هیچ مجوز خاصی برای مجموعه داده را فهرست نمی کند، اما مجوز آپاچی-2.0. وظیفه PF از مدل ها می خواهد که حوزه مطالعه یک مقاله را با توجه به عنوان آن پیش بینی کنند. مجموعه داده این وظیفه شامل 22,399/5,599/84,000 است

موسسه فناوری Massachu-setts (MIT) را فهرست می کند.

دو پروتئین در *GPT* برای این مجموعه داده مرتبط هستند، اگر و تنها در صورتی که در ادبیات علمی برای تعامل مستند شده باشند، با توجه به مجموعه داده های تعامل درخت زندگی. این یک نمودار دانش خارجی است.

برای FT، ما از وظایف ارزیابی تعبیه های پروتئین (TAPE) تسک های معیار FT[5]، از جمله همسانی از راه دور (RH)، یک دنباله استفاده می کنیم

فضای نهفته XX کمتر مورد بررسی قرار گرفته است (داده های توسعه یافته شکل 1). تحلیل های نظری و تجربی ما نشان می دهد که این کمبود مهم است. به طور مقطعی، ما یک چارچوب PT به نام SIPT را تعریف می کنیم که بر اساس آن از دست دادن PT به دو جزء تقسیم می شود: یکی که برای گرفتن روابط درون نمونه (به عنوان مثال، در هر نشانه) طراحی شده است و دیگری که برای محدود کردن فضای نهفته در هر نمونه برای ثبت رابطه بین نمونه های ارائه شده توسط GPT گراف PT مشخص شده توسط کاربر طراحی شده *است*. در چارچوب ما، ما از نظر تئوری و از طریق آزمایش نشان می دهیم که ساختار القا شده در XX می تواند مستقیما به عملکرد نهایی FT متصل شود. از نظر تجربی، ما نشان می دهیم که روش های SIPT با استفاده از انواع نمودارهای PT می توانند به طور مداوم از روش های PT موجود در سه حوزه دنیای واقعی بهتر عمل کنند.

کار ما چندین جهت مهم را برای تحقیقات آینده برجسته می کند. به عنوان مثال، آیا ضررهایی برای نمودارهای PT مناسب تر از ضررهای یادگیری متریک هستند - به عنوان مثال، آیا می توانیم از فاصله نمودار در کنار فاصله درون دسته ای برای بهبود استراتژی های نمونه گیری منفی استفاده کنیم؟ علاوه بر این، آیا می توانیم نتایج نظری در مورد همگرایی مدل های از پیش آموزش دیده ارائه دهیم؟ به عنوان مثال، آیا می توانیم درک زمان و نحوه همگرایی مدل های از پیش آموزش دیده را با راه حل هایی که *GPT* را بازیابی می کنند، پیش ببریم؟ در جهتی متفاوت، آیا مدل های از پیش آموزش دیده می توانند اشکال ساختار فراتر از روابط نزدیکترین همسایه را منعکس کنند - به عنوان مثال، با استفاده از ملاحظات توپولوژیکی مرتبه بالاتر یا با تطبیق یک تابع فاصله به جای یک نمودار گسسته؟ علاوه بر این، بررسی بیشتر تأثیر هدف القا کننده ساختار بر مکانیسم های داخلی مدل های زیربنایی، همانطور که از طریق تکنیک های هوش مصنوعی قابل توضیح بررسی می شود، راهی هیجان انگیز برای کارهای آینده خواهد بود. ما پیش بینی می کنیم که تجزیه و تحلیل بیشتر این سوالات و سایر سوالات منجر به روش های جدید PT شود و PT را قادر می سازد تا در حوزه های مختلف موفق باشد.

روش

زبان های القا کننده ساختار

ما از از دست دادن چند شباهت 46 استفاده می کنیم که با وزن جفت مثبت پارامتر می شود،

$+w$، وزن جفت منفی، $-w$، و هایپرپارامتر ثابت، T، که در زیر آورده شده است:

$$\mathrm{LSI} = \frac{1}{\text{شمال غربی}} \sum_{(i,j)\in E} \left(e^{-w_+\cdot((f_\theta(\mathbf{x}_i), f_\theta(\mathbf{x}_j))-t)} + \frac{1}{\text{شمال غربی}} \sum_{(i,j)\notin E} \left(e^{-w_-\cdot((f_\theta(\mathbf{x}_i), f_\theta(\mathbf{x}_j))-t)}\right)\right)$$

ورود به سیستم (1)

ورود به سیستم (1)

ما همچنین از یک ضرر تضادی که پس از نسخه در ref مدل شده است، استفاده می کنیم. 45. برای این از دست دادن، فرض می کنیم نگاشت های زیر به ما داده می شود: "pos"، که **x** را به یک گره مثبت نگاشت می کند (یعنی در GPT به x مرتبط است)، و "neg"، که **x** را به یک گره منفی نگاشت می کند (یعنی **در** GPT به x مرتبط نیست). اتحاد یک مینی دسته بذر B از نقاط **XB** و تصاویر آن تحت نگاشت های "pos" و "neg" یک مینی بچ کامل را تشکیل می دهد. این ضرر با پارامترهای حاشیه مثبت و منفی $+\mu$ و $-\mu$ به صورت زیر مشخص می شود:

$$L^{(CL)} = \frac{1}{N} \sum_{\mathbf{x}_i \in X} \left(-\log \frac{\exp(\mathrm{DD}(x_i, \mathrm{pos}(\mathbf{x}))/\tau)}{\exp(\mathrm{DD}(x_i, \mathrm{pos}(\mathbf{x}))/\tau) + \exp(\max(\mathrm{DD}(\mathbf{x}_i, \mathrm{neg}(\mathbf{x})), 0)/\tau)} \right)$$

$$\mathrm{DD}(x, y) = \max(\mu - \mathrm{DD}(\mathbf{x}, \mathrm{neg}(\mathbf{x})), 0).$$

مجموعه داده پروتئین ها و تسک های FT

ما از مجموعه داده ای از ~1.5 میلیون توالی پروتئینی از مجموعه داده درخت زندگی استنفورد[20] (https://snap.stanford.edu/tree-of-life/data.html) استفاده می کنیم. مخزن GitHub مرتبط با این منبع مجوز

https://doi.org/10.1038/s42256-023-00647-z

جملات قطار/اعتبارسنجی/تست اگرچه مجموعه داده اصلی از MAG[21] مشتق شده است، اما توسط SciBERT به طور مستقیم 53 در این قالب کار فرموله شده است. وظیفه SC مدل ها را برای پیش بینی برچسب "قصد" برای جملاتی که به سایر آثار علمی در مقالات دانشگاهی استناد می کنند، به چالش می کشد. مجموعه داده این وظیفه شامل 1,861/916/7,320 جمله قطار/اعتبارسنجی/تست است و در اصل از ref تهیه شده است. 54. وظیفه AA به مدل هایی نیاز دارد تا برچسب "قصد" را برای جملاتی که به سایر آثار علمی در مقالات دانشگاهی استناد می کنند، پیش بینی کنند. مجموعه داده این وظیفه شامل 139/114/1,688 جمله train/validation/test است و در اصل از ref تهیه شده است. 55.

ما با مدل منتشر شده SciBERT[53] به عنوان مقایسه هر نشانه و مدل BioLinkBERT[56] را به عنوان مقایسه هر نمونه مقایسه می کنیم. BioLinkBERT مدل سازی زبان را با یک وظیفه clas-sification تقویت می کند تا پیش بینی کند که آیا متن ورودی از دو سند از یک سند، اسناد پیوندی (که در آن پیوند وجود دارد) تشکیل شده است یا خیر

https://doi.org/10.1038/s42256-023-00647-z

شود. به این ترتیب، توجه داشته باشید که ما *λSI* را به گونه ای انتخاب می کنیم که inde-

از طریق نمودار استناد تعیین می شود) یا اسناد پیوند نشده. به این ترتیب، از اطلاعات مشابهی استفاده می کند که برای ساخت نمودار PT ما استفاده می شود، اما از طریق از دست دادن طبقه بندی تک وظیفه به جای تلفات کلی تر القا کننده ساختار که در اینجا استفاده می کنیم. اخیرا، مدل های زبان پایه موفق تری فراتر از مدل SciBERT (مانند PubMedBERT[57]) پیشنهاد شده اند و تغییر به استفاده از آنها برای مقداردهی اولیه مدل های SIPT ما در روش های شروع گرم احتمالا عملکرد را در همه مدل ها بهبود می بخشد. با این حال، با توجه به هزینه محاسباتی مدل PT، ما استفاده از SciBERT را برای مدل اولیه سازی خود حفظ می کنیم (و بر این اساس برای خط پایه هر توکن مربوطه) و بررسی PubMedBERT را برای کارهای آینده واگذار می کنیم.

مجموعه داده شبکه ها و تسک های FT

ما در اینجا از مجموعه داده ای از ~70,000 شبکه ایگو PPI استفاده می کنیم که از ref تهیه شده است.
26. هر نمونه در اینجا یک پروتئین واحد را توصیف می کند که به عنوان یک شبکه بیولوژیکی (یعنی یک نمودار منتسب) مربوط به شبکه ایگو در مورد آن پروتئین (یعنی یک زیرگراف کوچک حاوی تمام گره های پروتئین هدف) در یک نمودار PPI گسترده تر تحقق می یابد. برخلاف سایر دامنه های ما، این دامنه حاوی دنباله نیست. مجموعه داده Networks PT فایل های کد و مجموعه داده خود را تحت مجوز MIT منتشر می کند.

این مجموعه داده با حضور یا عدم وجود هر یک از 4,000 اصطلاح هستی شناسی ژن پروتئین مرتبط با پروتئین مرکزی در هر شبکه ایگو PPI برچسب گذاری شده است. با استفاده از این برچسب ها، دو شبکه ایگو PPI در GPT به هم متصل می شوند اگر و تنها در صورتی که فاصله همینگ بین بردارهای برچسب مشاهده شده آنها بیش از نه نباشد. این یک نمودار نزدیک ترین همسایه با نمایش جایگزین است.

ما فقط یک کار FT را در این محیط مطالعه می کنیم، که طبقه بندی باینری چند برچسبی 40 حاشیه نویسی اصطلاح Gene Ontology (متریک: منطقه ماکرو زیر منحنی مشخصه عملکرد گیرنده (AUROC)) است که در ref استفاده می شود. 26. ما از مجموعه PT برای آموزش FT استفاده می کنیم و مدل را بر روی یک تقسیم تصادفی 10٪ ارزیابی می کنیم.

ما با هر دو PT تحت نظارت 26 ویژگی و PT تحت نظارت چند وظیفه مقایسه می کنیم.

راه اندازی آزمایشی

برای به حداقل رساندن بار محاسباتی، ما یک مدل القا کننده ساختار را از ابتدا برای مجموعه داده های پروتئین ها و چکیده ها از قبل آموزش نمی دهیم. در عوض، ما یک مدل را مستقیما از خط پایه هر توکن مقداردهی اولیه می کنیم ، سپس PT اضافی را تنها برای تعداد کمی از دوره ها تحت زیرمجموعه ضرر SIPT انجام می دهیم. ما هر دو واریانت LSI چند شباهت و کنتراست را در این حوزه ها ارزیابی می کنیم. در مجموعه داده Networks، ما همه مدل ها (از جمله خطوط پایه) را از ابتدا آموزش می دهیم و بر اساس نتایج تجربی اولیه، فقط نوع ضرر تضادی را ارزیابی می کنیم.

تجزیه و تحلیل فرسایش

توجه داشته باشید که روش شروع گرم که در بالا در حوزه های پروتئین ها و چکیده ها توضیح داده شد، امکان یک مطالعه فرسایش قدرتمند را فراهم می کند: علاوه بر این، با آموزش یک مدل PT از خط پایه هر توکن با $\lambda SI = 0$، می توانیم به طور منحصر به فرد تأثیر اصطلاح ضرر جدید را ارزیابی کنیم، به جای آموزش اضافی یا مجموعه داده های مختلف PT. ما این مطالعه فرسایش را برای همه مجموعه داده های مرتبط انجام می دهیم. برای مجموعه داده های Networks، با توجه به اینکه همه مدل ها از ابتدا با همان روش های توقف زودهنگام آموزش دیده اند، نیازی به مطالعه دیگری برای ارزیابی تأثیر مدت از دست دادن نیست.

انتخاب *پارامتر مدل* λSI

برای مجموعه داده های پروتئین ها و چکیده ها، برای انتخاب مقدار بهینه *λSI* برای استفاده در زمان PT، چندین مدل را از قبل آموزش دادیم و اثربخشی آنها را در یک کار بازیابی پیوند بر روی $GPT = (V, E)$ ارزیابی کردیم. به طور خاص، ما با جاسازی همه گره *های* $n \in V$ به عنوان $f(n)$ یک گره جاسازی *شده را به دست* می آوریم، سپس تمام گره های دیگر n' را با فاصله اقلیدسی بین $f(n)$ و $f(n')$ رتبه بندی می کنیم، و این لیست رتبه بندی شده را از طریق دقت میانگین رتبه بندی برچسب، سود تجمعی تخفیف نرمال شده، دقت متوسط و میانگین رتبه متقابل ارزیابی کنید، جایی که یک گره n' به عنوان یک بازیابی "موفق" برای $(n, n') \in E$ if n *در نظر گرفته می*

https://doi.org/10.1038/s42256-023-00647-z

معلق وظیفه FT است و می تواند صرفا بر اساس داده های PT تعیین شود. نتایج نهایی برای این آزمایش ها در جدول داده های توسعه یافته 5 برای مجموعه داده پروتئین ها و جدول داده های توسعه یافته 6 برای مقالات علمی نشان داده شده است. در نهایت، این فرآیند نشان می دهد که 0.1 λSI یک تنظیم قوی است و به این ترتیب، 0.1 به طور مستقیم برای وظیفه شبکه ها بدون بهینه سازی بیشتر استفاده شده است.

معماری مدل و سایر پارامترهای مدل

معماری رمزگذارهای ما برای حوزه های پروتئین ها و چکیده ها به طور کامل از مدل های منبع ما در TAPE[5] و SciBERT[53] تعیین شده است. به طور خاص، برای پروتئین ها و مقالات علمی، ما از یک ترانسفورماتور 12 لایه با اندازه پنهان 768، اندازه متوسط 3072 و 12 سر توجه استفاده می کنیم. توکنایزرهای ارائه شده TAPE و SciBERT نیز استفاده می شود. یک لایه خطی واحد به ابعاد خروجی هر کار به عنوان سر پیش بینی استفاده می شود و خروجی توکن [CLS] لایه نهایی را به عنوان یک جاسازی دنباله کامل به عنوان ورودی در نظر می گیرد. ما همچنین PT را برای یک یا چهار دوره اضافی بر اساس عملکرد مجموعه اعتبارسنجی آزمایش کردیم. ما در نهایت از یک دوره برای پروتئین ها و چهار دوره برای مقالات علمی استفاده کردیم.

برای دامنه شبکه ها، ما معماری مورد استفاده در منبع اصلی[26] را برای اجرای مدل ماسک مطابقت می دهیم، آن را برای کارایی محاسباتی ذخیره می کنیم، اندازه دسته را تا حد امکان افزایش می دهیم، سپس به تناسب نرخ یادگیری را افزایش می دهیم تا اندازه دسته بزرگتر را در نظر بگیریم. این مربوط به اندازه دسته ای 1024، نرخ یادگیری 0.01، یک شبکه عصبی انقلابی گراف (GCNN) با رمزگذار شبکه ایزومورفیسم گراف (GIN)، تعبیه ابعاد 300، 5 لایه، 10٪ افت، تجمیع میانگین و استراتژی ترکیب ویژگی گره (JK) "آخرین" است.

فراپارامترهای FT (نرخ یادگیری، اندازه دسته و تعداد دوره ها) بر اساس ترکیبی از نتایج موجود، تنظیم فراپارامترها و محدودیت های ماشین تعیین شد. در پروتئین ها، اکثر ابرپارامترها به گونه ای تنظیم شدند که از موارد گزارش شده برای مدل LM PT در ref پیروی کنند. 58، اگرچه جستجوهای هایپرپارامتر محدود اضافی برای تأیید کافی بودن این انتخاب ها انجام شد. از آنجایی که منبع اصلی این فراپارامترها یک مدل LM PT بود، هر گونه تعصب در اینجا باید علیه SIPT باشد، به این معنی که این یک انتخاب محافظه کارانه است. توقف زودهنگام (بر اساس تعداد دوره ها بدون مشاهده بهبود در عملکرد مجموعه اعتبارسنجی) استفاده شد و اندازه دسته تا حد امکان با در نظر گرفتن ماشین زیرین تنظیم شد. برای بازتولید PLUS، فراپارامترهای مشابه با ابرپارامترهای PLUS گزارش شده را برای سایر وظایف و مشابه فراپارامترهای خود برای سایر وظایف مقایسه کردیم و از پارامترهایی استفاده کردیم که بهترین عملکرد را در مجموعه اعتبارسنجی داشتند. برای مقالات علمی، ما یک جستجوی شبکه ای را برای بهینه سازی عملکرد وظایف پایین دستی در مجموعه اعتبارسنجی انجام دادیم، با نرخ یادگیری بین 5 × 10-6 و 5 × 10-5 و تعداد دوره ها بین 2 تا 5. همان جستجوی شبکه ای در روش اصلی SciBERT استفاده شد. علاوه بر این، ما با استفاده از بهینه ساز Adam با گرم کردن و پوسیدگی خطی، اندازه دسته ای 32 و بدون توقف زودهنگام، معیار SciBERT را مطابقت می دهیم. برای شبکه ها، فراپارامترهای FT دوباره برای مطابقت با مدل منبع اصلی 26 انتخاب شدند تا از افزایش اندازه دسته ای و نرخ یادگیری صرفه جویی شود. هیچ جست وجوی فراپارامتر اضافی انجام نشد.

فراپارامترهای نهایی برای هر کار پایین دستی در جدول داده های توسعه یافته 1 برای پروتئین ها و جدول داده های توسعه یافته 2 برای مقالات علمی نشان داده شده است.

پیاده سازی و محاسبات محیط

ما از PyTorch برای پایگاه کد خود استفاده می کنیم. FT Experiments و Networks PT بر روی ماشین های مختلف اوبونتو (نسخه های مختلف از 16.04 تا 20.04) با واحدهای پردازش گرافیکی مختلف NVIDIA اجرا شد. پروتئین ها و چکیده ها اجرای PT بر روی یک سیستم Power 9 انجام شد که هر کدام با استفاده از 4 واحد پردازش گرافیکی NVIDIA 32 گیگابایتی V100 با InfiniBand با نصف دقت اجرا شدند.

مرور سیستماتیک روش های PT

مقالات از طریق جست وجوی دستی در NLP و NLP مشتق شده انتخاب شدند

روش های PT (یعنی روش هایی که عمدتا بر حوزه های دیگر یا دامنه های چند وجهی متمرکز شده اند) از طریق Google Scholar و با خزیدن از طریق منابع مقالاتی که قبلا گنجانده شده اند، حذف شدند. تعداد استنادات برای هر اثر از طریق Google Scholar در 2 آگوست 2022 به دست آمد. تاریخ انتشار (که برای محاسبه استنادها در ماه از تاریخ انتشار استفاده می شود) به عنوان زودتر از (1) تاریخ انتشار خاص محل مقاله یا (2) اولین تاریخ ارسال به پلتفرم های arXiv یا bioRxiv، همانطور که از طریق تطابق دقیق عنوان ارجاع داده شده است، محاسبه شد. یک بررسی دستی برای طبقه بندی نحوه محدود کردن روش های PT هندسه فضای پنهان و اختصاص امتیازات محورهای ذهنی، عددی "کم عمق-عمیق" و "صریح-ضمنی" انجام شد. در مجموع، بیش از 90 روش مورد بررسی قرار گرفت که 74 مورد از آنها برای گنجاندن در نتایج بررسی عددی مناسب بودند (شکل داده های توسعه یافته 1). اطلاعات تکمیلی تمام روش های در نظر گرفته شده را خلاصه و دسته بندی می کند (و دلایل خروج آورده شده است). توجه داشته باشید که چارچوب ما بر روی روش های PT مشتق شده از NLP تمرکز دارد، اما ما روش PT مولد متمرکز بر توزیع های پیوسته با ابعاد بالا، مانند مدل های انتشار[59] را بررسی نمی کنیم. با این حال، این روش ها در حوزه های دیگر مانند بینایی کامپیوتر موفق بوده اند.

در دسترس بودن داده ها

مجموعه داده های مصنوعی و اشاره گرهای ما به مجموعه داده های دنیای واقعی در آموزش https://github.com/mmcdermott/structure_inducing_pre به صورت عمومی در دسترس هستند .

در دسترس بودن کد

پیاده سازی روش توسعه یافته و مورد استفاده در مطالعه در پایتون از طریق وب سایت پروژه به آدرس https://zitniklab.hms.harvard.edu/projects/SIPT در دسترس است. کد بازتولید نتایج، مستند سازی و نمونه های استفاده در آموزش https://github.com/mmcdermott/structure_inducing_pre است.

مراجع

1. Devlin، J.، Chang، M.-W.، Lee، K. & Toutanova، K. BERT: پیش آموزش ترانسفورماتورهای دو طرفه عمیق برای درک زبان. در *مجموعه مقالات کنفرانس 2019 بخش آمریکای شمالی انجمن زبان شناسی محاسباتی: فناوری های زبان انسانی، جلد 1 (مقالات بلند و کوتاه)* 4171-4186 (انجمن زبان شناسی محاسباتی، 2019).
2. Deng، J. et al. Imagenet: یک پایگاه داده تصویر سلسله مراتبی در مقیاس بزرگ . در *سال 2009 کنفرانس IEEE در مورد بینایی کامپیوتر و تشخیص الگو* 248-255 (2009 ، IEEE).
3. براون، تی بی و همکاران مدل های زبانی یادگیرندگان کمی هستند. در *مجموعه مقالات سی و چهارمین کنفرانس بین المللی سیستم های پردازش اطلاعات عصبی* 1877،**33**- 1901 (NIPS، 2020).
4. Sanh، V. و همکاران آموزش چند وظیفه ای تعمیم وظایف بدون شات را امکان پذیر می کند. در *کنفرانس بین المللی بازنمایی های یادگیری* (2022).
5. Rao، R. و همکاران ارزیابی یادگیری انتقال پروتئین با TAPE. در *پیشرفت در سیستم های پردازش اطلاعات عصبی جلد 32* (ویراستاران والاچ ، اچ و همکاران) (Curran Associates ، 2019).
6. Schwaller، P.، Hoover، B.، Reymond، Jean-Louis، Strobelt، H. & Laino، T. استخراج دستور زبان شیمی آلی از یادگیری بدون نظارت واکنش های شیمیایی. *علمی وکالت* **7** ، (2021) eabe4166.
7. لی ، ب. و همکاران. در جاسازی های جمله از مدل های زبان از پیش آموزش دیده. در *Proc. کنفرانس 2020 روش های تجربی در پردازش زبان طبیعی* 9119-9130 (انجمن زبان شناسی محاسباتی ، 2020).
8. Liu، Y. و همکاران RoBERTa: یک رویکرد پیش آموزش BERT به شدت بهینه شده. پیش چاپ در (2019) https://arxiv.org/abs/1907.11692.
9. Radford، A.، Narasimhan، K.، Salimans، T. & Sutskever، I. بهبود درک زبان با پیش آموزش مولد. (2018).
10. Lan، Z. و همکاران آلبرت: یک BERT ساده برای یادگیری خودنظارتی بازنمایی های زبانی. در *کنفرانس بین المللی بازنمایی های یادگیری* (ICLR، 2019).
11. Liu، X.، He، P.، Chen، W. & Gao، J. شبکه های عصبی عمیق چند وظیفه ای برای درک زبان طبیعی. در *مجموعه مقالات پنجاه و هفتمین نشست سالانه انجمن زبان شناسی محاسباتی* (ویراستاران Korhonen، A. et al.) 4487-4496 (ACL، 2019).
12. Giorgi، J.، Nitski، O.، Wang، B. & Bader، G. DeCLUTR: یادگیری تضاد عمیق برای بازنمایی های متنی بدون نظارت. در *پنجاه و نهمین نشست سالانه انجمن زبان شناسی محاسباتی و یازدهمین کنفرانس مشترک بین المللی پردازش زبان طبیعی* جلد 1، 879-895 (انجمن زبان شناسی محاسباتی، 2021).
13. کنگ ، L. و همکاران. دیدگاه بیشینه سازی اطلاعات متقابل یادگیری بازنمایی زبان. در *کنفرانس بین المللی بازنمایی های یادگیری* (2020).
14. Tetko, I. V., Karpov, P., Bruno, E., Kimber, T. B. & Godin, G. تقویت چیزی است که شما نیاز دارید! در *کنفرانس بین المللی شبکه های عصبی مصنوعی* 831-835 (اسپرینگر ، 2019).
15. Karpov، P.، Godin، G. & Tetko، I. V. Transformer-CNN: چاقوی سوئیسی برای مدل سازی و تفسیر QSAR. *J. Cheminform.* 12, 1-12 (2020).
16. Tetko، I. V.، Karpov، P.، Van Deursen، R. & Godin، G. مدل های ترانسفورماتور NLP تقویت شده پیشرفته برای رتروسنتز مستقیم و تک مرحله ای. *Nat. Commun.* 11, 1-11 (2020).
17. Schwaller، P. et al. ترانسفورماتور مولکولی: مدلی برای پیش بینی واکنش شیمیایی کالیبره شده با عدم قطعیت. *سنت* 5, 1572-1583 (2019) *ACS. Sci.*.
18. Wu، Z. و همکاران CLEAR: یادگیری تقابلی برای بازنمایی جمله. پیش چاپ در (2020) https://arxiv.org/abs/2012.15466.
19. منگ، ی. و همکاران COCO-LM: تصحیح و تضاد توالی های متنی برای پیش آموزش مدل زبان. در *Adv. Neural Inf. روند. سیستم.* (ویراستاران رانزاتو، م. و همکاران.) **34**، 23102-23114 (Curran Associates، 2021).
20. Zitnik، M.، Sosič، R.، Feldman، M. W. & Leskovec، J. تکامل انعطاف پذیری در تعاملات پروتئینی در سراسر درخت زندگی. *Proc. Natl Acad. علمی ایالات متحده آمریکا* **116** ، 4426-4433 (2019).
21. وانگ ، ک. و همکاران. مروری بر خدمات آکادمیک مایکروسافت برای مطالعات علوم علوم. *جلو. کلان داده* **2** (2019).
22. Hu، W. و همکاران معیار نمودار باز: مجموعه داده ها برای یادگیری ماشین در نمودارها. در *پیشرفت در سیستم های پردازش اطلاعات عصبی* **33** ، 22118-22133 (NEURIPS ، 2020).
23. Fang، Y. et al. یادگیری نمودار مولکولی کنتراستیو آگاه از دانش. پیش چاپ در (2021) https://arxiv.org/abs/2103.13047.
24. Sanh، V. و همکاران آموزش چند وظیفه ای تعمیم وظایف بدون شات را امکان پذیر می کند. در *کنفرانس بین المللی بازنمایی های یادگیری* (2021).
25. Rives، A. et al. ساختار و عملکرد بیولوژیکی از مقیاس بندی یادگیری بدون نظارت تا 250 میلیون توالی پروتئینی پدیدار می شود.*Proc. Natl Acad. علمی ایالات متحده آمریکا* **118** ، (2021) e2016239118.
26. هو ، دبلیو و همکاران. استراتژی های پیش آموزش شبکه های عصبی گراف. در *ICLR* (2020).
27. مک درموت ، M. B. A. و همکاران. یک معیار جامع سری زمانی قبل از آموزش او. در *مجموعه مقالات کنفرانس سلامت، استنباط و یادگیری، CHIL '21* 257-278 (ACM، 2021).
28. رائو ، آر ام و همکاران ترانسفورماتور MSA. در *Proc. سی و هشتمین کنفرانس بین المللی یادگیری ماشین، Proc. تحقیقات یادگیری ماشین* ، جلد 139 (ویراستاران Meila، M. & Zhang، T.) 8844-8856 (PMLR، 2021).
29. Saunshi, N., Plevrakis, O., Arora, S., Khodak, M. & Khandeparkar, H. تحلیل نظری یادگیری بازنمایی بدون نظارت تضادی. در *مجموعه مقالات سی و ششمین کنفرانس بین المللی یادگیری ماشین، جلد 97* (ویراستاران Chaudhuri، K. & Salakhutdinov، R.) 5628-5637 (PMLR، 2019).
30. لوین ، ی. و همکاران. (1401). سوگیری استقرایی یادگیری درون زمینه: بازاندیشی در طراحی مثال پیش آموزش. در *کنفرانس بین المللی بازنمایی های یادگیری* (2022).

50. Sarkisyan، K. S. et al. چشم انداز تناسب اندام محلی پروتئین فلورسنت سبز. *طبیعت* **533**, 397-401 (2016).
51. AlQuraishi، M. ProteinNet: یک مجموعه داده استاندارد برای یادگیری ماشین ساختار پروتئین. (2019) 20, 1-10 *BMC Bioinform.*.
52. Min، S.، Park، S.، Kim، S.، Choi، H.-S. & Yoon، S پروتئین دو طرفه عمیق با اطلاعات ساختاری. *دسترسی* -123912 9، *IEEE* 123926 (2021).
53. Beltagy، I.، Lo، K. & Cohan، A. SciBERT: یک مدل زبانی از پیش آموزش دیده برای متن علمی. در *مجموعه مقالات کنفرانس 2019 روش های تجربی در پردازش زبان طبیعی و نهمین کنفرانس بین المللی مشترک پردازش زبان طبیعی (EMNLP-IJCNLP)* 3615-3620 (ACL، 2019).
54. Cohan، A.، Ammar، W.، van Zuylen، M. & Cady، F. داربست های ساختاری برای طبقه بندی قصد استناد در نشریات علمی. در *مجموعه مقالات کنفرانس 2019 بخش آمریکای شمالی انجمن زبان شناسی محاسباتی: فناوری های زبان انسانی، جلد 1 (مقالات بلند و کوتاه)* 3586–3596 (ACL، 2019).
55. Jurgens، D.، Kumar، S.، Hoover، R.، McFarland، D. & Jurafsky، D. اندازه گیری تکامل یک رشته علمی از طریق چارچوب های استناد. ترجمه *Assoc. محاسبات. زبان شناس.* **6**, 391-406 (2018).
56. Yasunaga، M.، Leskovec، J. & Liang، P. LinkBERT: مدل های زبان پیش آموزش با پیوندهای سند. در *Proc. شصتمین نشست سالانه انجمن زبان شناسی محاسباتی* جلد 1، 8003-8016 (انجمن زبان شناسی محاسباتی، 2022).
57. Gu، Y. و همکاران مدل زبان خاص دامنه پیش آموزش برای پردازش زبان طبیعی زیست پزشکی. *ACM Trans. Computi. سلامت*(2021) 3, 1-23 *c.*.
58. مک درموت ، م. ، یاپ ، ب. ، هسو ، اچ ، جین ، دی و سولوویتس ، پ. خصمانه پیش آموزش کنتراستیکی برای توالی های پروتئینی. پیش چاپ در https://arxiv.org/abs/2102.00466 (2021).
59. رامش ، A. ، Dhariwal ، P. ، Nichol ، A. ، Chu ، C. & Chen ، M. تولید تصویر شرطی متن سلسله مراتبی با نهفته کلیپ. پیش چاپ در https://arxiv.org/abs/2204.06125 (2022).

تقدیر و تشکر

MBAM تا حدی توسط LM013337 کمک هزینه مؤسسه ملی بهداشت (NIH) و یک توافق نامه تحقیقاتی مشترک با IBM پشتیبانی شد.
و همچنین توسط دانشکده پزشکی هاروارد گروه انفورماتیک زیست پزشکی برکوویتز فلوشیپ فوق دکتری. BY توسط صندوق فرصت های تحقیقاتی در مقطع کارشناسی موسسه فناوری ماساچوست (MIT) پشتیبانی شد. M.Z. با سپاسگزاری از حمایت NIH R01HD108794، شماره قرارداد نیروی هوایی ایالات متحده قدردانی می کند. FA8702-15-D-0001، و جوایزی از ابتکار علوم داده هاروارد، تحقیقات دانشکده آمازون، برنامه Google Research Scholar، Bayer Early Excellence in Science، AstraZeneca Research و Roche Alliance with Universities برجسته. هر گونه نظرات، یافته ها، نتیجه گیری ها یا توصیه های بیان شده در این مطالب متعلق به نویسندگان است و لزوما منعکس کننده دیدگاه های سرمایه گذاران نیست.

مشارکت های نویسنده

M.B.A.M. و BY مجموعه داده ها را جمع آوری کردند، کد مدل سازی را نوشتند و آزمایش هایی را انجام دادند. M.B.A.M. نتایج نهایی را گردآوری کرد و بررسی مطالعات پیش از آموزش موجود را تکمیل کرد. M.B.A.M.، P.S و M.Z.
مطالعه را تصور کرد و چارچوب کار را شکل داد. PS و M.Z. بینش و راهنمایی را در طول پروژه ارائه دادند. M.B.A.M. و M.Z. مقاله نهایی را نوشت و P.S، B.Y.، M.B.A.M. و M.Z. در ویرایش پیش نویس ها مشارکت داشتند.

تعارض منافع

نویسندگان هیچ تعارض منافع را اعلام نمی کنند.

اطلاعات اضافی:

داده های توسعه یافته برای این مقاله در https://doi.org/10.1038/s42256-023-00647-z در دسترس است.

31. Radford، A. et al. مدل های زبانی یادگیرندگان چند وظیفه ای بدون نظارت هستند. (2019) 9 ،1 *OpenAI*.
32. Ribeiro، D. N. & Forbus، K. ترکیب قیاس با مدل های زبانی برای استخراج دانش. در *سومین کنفرانس ساخت خودکار پایگاه دانش* (1400).
33. Wang، X. et al. KEPLER: یک مدل یکپارچه برای جاسازی دانش و بازنمایی زبان از پیش آموزش دیده. *ترجمه Assoc. محاسبات. زبان شناس.* **9**، 176-194 (2021).
34. Jiang، X.، Liang، Y.، Chen، W. & Duan، N. XLM-K: بهبود پیش آموزش مدل زبان بین زبانی با چند زبانه دانش. در *Proc. کنفرانس هوش مصنوعی AAAI* جلد 36, 10840–10848 (2022).
35. Guo، Y. et al. Webformer: پیش آموزش با صفحات وب برای بازیابی اطلاعات. در *Proc. چهل و پنجمین کنفرانس بین المللی ACM SIGIR در مورد تحقیق و توسعه در بازیابی اطلاعات* 1502-1512 (ACM، 2022).
36. گائو، H. و هوانگ، H. تعبیه شبکه عمیق منسوب. در *فرآیندهای بیست و هفتمین کنفرانس مشترک بین المللی هوش مصنوعی* 3364-3370 (انتشارات AAAI ، 2018).
37. Cui، G.، Zhou، J.، Yang، C. & Liu، Z. رمزگذار گراف تطبیقی برای جاسازی نمودار نسبت داده شده. در *Proc. 26th ACM SIGKDD کنفرانس بین المللی کشف دانش و داده کاوی* 976–985 (ACM، 2020).
38. Li، Y.، Sha، C.، Huang، X. & Zhang، Y. تشخیص جامعه در نمودارهای منتسب شده: یک رویکرد جاسازی. در *Proc. کنفرانس هوش مصنوعی AAAI* جلد 32 (2018).
39. Li، M. M.، Huang، K. & Zitnik، M. یادگیری نمایندگی برای شبکه های زیست شناسی و پزشکی: پیشرفت ها، چالش ها و فرصت ها. پیش چاپ در https://arxiv.org/abs/2104.04883 (2021).
40. Kipf، T. N. & Welling، M. طبقه بندی نیمه نظارت شده با شبکه های کانولوشنی گراف. در *پنجمین کنفرانس بین المللی بازنمایی های یادگیری* (OpenReview.net، 2017).
41. Hamilton، W.، Ying، Z. & Leskovec، J. یادگیری بازنمایی استقرایی در نمودارهای بزرگ. در *پیشرفت در سیستم های پردازش اطلاعات عصبی ، جلد 30* (ویراستاران Guyon، I. و همکاران) 1025-1035 (2017).
42. Vert، J.-P. & Yamanishi، Y. استنتاج گراف تحت نظارت. در *پیشرفت در سیستم های پردازش اطلاعات عصبی ، جلد 17* (ویراستاران L ، Saul. و همکاران) (MIT Press ، 2004).
43. Shaw, B. & Jebara, T. ساختار حفظ جاسازی در *مجموعه مقالات بیست و ششمین کنفرانس بین المللی سالانه یادگیری ماشین، ICML '09* (ACM، 2009).
44. Shaw، B.، Huang، B. & Jebara، T. یادگیری یک متریک از راه دور از یک شبکه. در *پیشرفت در سیستم های پردازش اطلاعات عصبی ، جلد 24* (ویراستاران J ، Shawe-Taylor. و همکاران) (Curran Associates ، 2011).
45. Hadsell، R.، Chopra، S. & LeCun، Y. کاهش ابعاد با یادگیری نقشه برداری ثابت. در *سال IEEE Computer 2006 کنفرانس جامعه بینایی کامپیوتر و تشخیص الگو (CVPR'06) ، جلد 2.* 1735–1742 (2006).
46. وانگ ، ایکس ، هان ، ایکس ، هوانگ ، دبلیو ، دونگ ، دی و اسکات ، م. آر. از دست دادن چند شباهت با وزن دهی جفتی عمومی برای یادگیری عمیق متریک. در *کنفرانس IEEE در مورد بینایی کامپیوتر و تشخیص الگو، CVPR 2019* 5022–5030 (بنیاد بینایی کامپیوتر/ IEEE، 2019).
47. Hou، J.، Adhikari، B. & Cheng، J. DeepSF: شبکه عصبی کانولوشن عمیق برای نقشه برداری از توالی های پروتئینی به چین ها. *بیوانفورماتیک* **34** ، 1295-1303 (2018).
48. Klausen، M. S. et al. NetSurfP-2.0: بهبود پیش بینی ویژگی های ساختاری پروتئین با یادگیری عمیق یکپارچه. *پروتئین ها* **87** ، 520-527 (2019).
49. Rocklin، G. J. et al. تجزیه و تحلیل جهانی تاشو پروتئین با استفاده از طراحی، سنتز و آزمایش موازی انبوه. *علوم پایه* **357** ، 168-175 (2017).

اقتباس، توزیع و تکثیر در هر رسانه یا قالب، تا زمانی که اعتبار مناسبی به نویسنده(های) اصلی و منبع بدهید، پیوندی به مجوز Creative Commons ارائه دهید و مشخص کنید که آیا تغییراتی ایجاد شده است یا خیر. تصاویر یا سایر مطالب شخص ثالث در این مقاله در مجوز Creative Commons مقاله گنجانده شده است، مگر اینکه در خط اعتباری مطالب خلاف آن ذکر شده باشد. اگر مطالبی در مجوز Creative Commons مقاله گنجانده نشده باشد و استفاده مورد نظر شما توسط مقررات قانونی مجاز نباشد یا از استفاده مجاز فراتر رود، باید مستقیما از دارنده حق نسخه برداری مجوز بگیرید.
برای مشاهده یک کپی از این مجوز، به http://creativecommons.org/licenses/by/4.0/ مراجعه کنید.

© نویسندگان: 1402

اطلاعات تکمیلی نسخه آنلاین حاوی مطالب تکمیلی موجود در https://doi.org/10.1038/s42256-023-00647-z است.

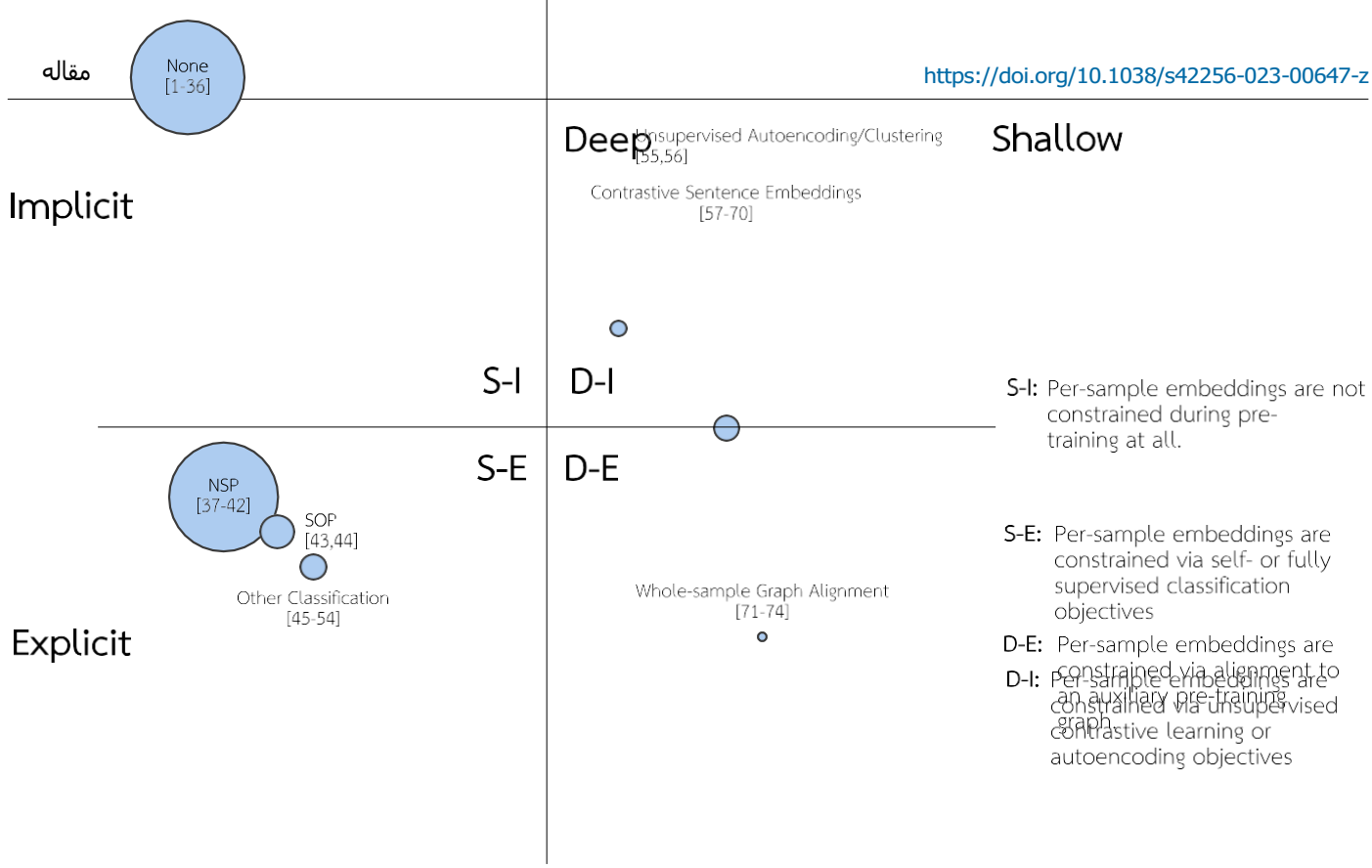
مکاتبات و درخواست های مواد باید به مارینکا زیتنیک ارسال شود.

اطلاعات تجدید چاپ و مجوزها در www.nature.com/reprints موجود است.

یادداشت ناشر: Springer Nature در رابطه با ادعاهای قضایی در نقشه های منتشر شده و وابستگی های نهادی بی طرف باقی می ماند.

دسترسی آزاد این مقاله تحت مجوز Creative Commons Attribution 4.0 International License است که اجازه استفاده، اشتراک گذاری و

https://doi.org/10.1038/s42256-023-00647-z



داده های توسعه یافته شکل 1 | روش های موجود پیش آموزش (PT).
خلاصه ای از 74 روش پردازش زبان طبیعی (NLP) و PT مشتق شده از NLP، بر اساس نحوه اعمال محدودیت های ساختاری به خوشه ها طبقه بندی می شوند
روی فضای نهفته PT (در هر نمونه). خوشه ها از طریق قضاوت دستی در مورد اینکه آیا محدودیت تحمیل *شده کم عمق* در مقابل *عمیق* و *ضمنی* در مقابل *صریح* است، بر روی محورها مرتب می شوند. اندازه خوشه ها به گونه ای است که مساحت مربوط به تعداد باشد
از روش های استنادی موجود در آن خوشه به طور متوسط در هر ماه از زمان اولین انتشار، با توجه به تعداد استنادات Google Scholar دریافت کرده اند.

"هیچکدام" مدل هایی را ثبت می کند که از ضرر قبل از آموزش نسبت به جاسازی هر نمونه استفاده نمی کنند. "NSP" به "پیش بینی جمله بعدی" اشاره دارد، وظیفه PT به ازای هر نمونه معرفی شده در "BERT[1]. "SOP" به "پیش بینی ترتیب جمله" اشاره دارد، وظیفه PT برای هر نمونه معرفی شده در ALBERT[10]. توجه داشته باشید که در مجموع بیش از 90 مطالعه در مرور ما در نظر گرفته شدند، اما فقط 74 مطالعه معیارهای ورود را برای ورود به این رقم داشتند. این روش ها با جزئیات بیشتری در اطلاعات تکمیلی شرح داده شده است.

https://doi.org/10.1038/s42256-023-00647-z

جدول داده های توسعه یافته **1** | فراپارامترهای نهایی برای دامنه پروتئین ها

Lr	اندازه دسته ای	وظیفه
1E-5	16	همسانی از راه دور
5e-5	128	فلورسانس
1e-4	512	ثبات
1E-5	16	ساختار ثانویه

فراپارامترهای نهایی برای حوزه پروتئین های ما. همه وظایف از 200 دوره در مجموع استفاده کردند و پس از 25 دوره بدون بهبود مجموعه اعتبارسنجی، توقف زودهنگام انجام شدند. LR، نرخ یادگیری.

https://doi.org/10.1038/s42256-023-00647-z

جدول داده های توسعه یافته **2 |** فراپارامترهای نهایی برای مجموعه داده **UMBRELLAS**

Lr	تعداد دوره ها	وظیفه
5e-5	2	زمینه مقاله
5e-5	5/4	ACL-ARC
1E-5	2/3	SciCite

فراپارامترهای نهایی برای مجموعه داده چکیده های ما. همه مدل ها از اندازه دسته ای 32 و بدون توقف زودهنگام برای مطابقت با کاغذ اصلی SciBERT[53] استفاده می کردند. LR، نرخ یادگیری. A / B = [هایپرپارامتر LM PT] / [هایپرپارامتر SIPT].

https://doi.org/10.1038/s42256-023-00647-z

جدول داده های توسعه یافته **3** | نتایج برای حوزه پروتئین ها

مدل	Rh	Fl	خیابان	Ss	Cp
نوار	%21	**0.68**	0.73	%73	0.32
پلاس	٪19.8 ± 1.7*	0.63	0.76	%73	N / A
ال ام پی تی	٪23.8 ± 1.1	0.00 ± 0.67	0.02 ± 0.76	٪73.9 ± 0.0	0.38
سیپت-سی	٪25.1 ± 0.6	**0.00 ± 0.68**	**0.01 ± 0.77**	٪73.9 ± 0.0	0.38
SIPT-M	**٪26.6 ± 1.0**	**0.00 ± 0.68**	0.01 ± 0.76	**٪74.2 ± 0.1**	**0.39**

نتایج ترانسفورماتور TAPE[5] ، ترانسفورماتور PLUS[52] (*: اندازه گیری های ما) ، خط پایه LM PT و دو نوع SIPT ("-C" نشان دهنده از دست دادن کنتراستیکی ، "-M" از دست دادن چند شباهت است). بالاتر بهتر است و بهترین نتایج در هر کار پررنگ هستند.

https://doi.org/10.1038/s42256-023-00647-z

جدول داده های توسعه یافته **4** | نتایج برای دامنه چکیده ها

مدل	Pf	Sc	Aa	SRE
SciBERT	**0.66**	0.85	0.71	0.80
BioLinkBERT	**0.0 ± 0.66**	**0.01 ± 0.86**	0.04 ± 0.73	**0.02 ± 0.82**
ال ام پی تی	**0.0 ± 0.66**	0.01 ± 0.85	0.05 ± 0.70	0.01 ± 0.80
سیپت-سی	**0.0 ± 0.66**	**0.01 ± 0.86**	**0.02 ± 0.76**	0.00 ± 0.81
SIPT-M	**0.0 ± 0.66**	0.00 ± 0.85	0.05 ± 0.73	N / A

نتایج مدل اصلی SciBERT[53] ، خط پایه LM PT خودمان و دو نوع SIPT ("C-" نشان دهنده از دست دادن کنتراست، "M-" از دست دادن چند شباهت). بالاتر بهتر است و بهترین نتایج در هر کار پررنگ هستند.

https://doi.org/10.1038/s42256-023-00647-z

جدول داده های توسعه یافته **5 |** عملکرد بازیابی پیوند **PT** برای حوزه پروتئین ها

روش	*λSI*	LRAP	nDCG	Ap	MRR
خط پایه تصادفی	N / A	%0.88	%27.1	%0.88	0.003
نوار[5]	N / A	%8.50	%34.9	%2.41	0.226
LM PT خط پایه	0	%8.92	%38.0	%2.33	0.238
SIPT (TAPE مقداردهی اولیه)	0.01	%9.69	%39.1	%2.56	0.254
	0.10	%10.95	%39.4	%3.46	0.260
	0.50	%10.54	%40.3	%3.43	0.246
	0.90	%10.12	%39.0	%3.16	0.237
	0.99	%14.50	%37.5	%3.13	0.236

PT عملکرد بازیابی پیوند را برای یک خط پایه تصادفی، مدل TAPE خام و SIPT برای پارامترهای مختلف وزن دهی *λSI* بر روی مجموعه داده های توالی پروتئینی تنظیم کرد. LRAP، میانگین دقت رتبه بندی برچسب. nDCG، سود تجمعی با تخفیف نرمال شده. AP، دقت متوسط؛ MRR، میانگین رتبه متقابل. مقادیر بالاتر نشان دهنده عملکرد بهتر است. با رنگ خاکستری برجسته شده است تحقق چارچوب SIPT که نتایج بهتری نسبت به قوی ترین خط پایه به همراه دارد و شواهدی را ارائه می دهد که ترکیب اطلاعات رابطه ای در سطح توالی در PT (یعنی $\lambda SI > 0$) منجر به بهبود عملکرد می شود.

https://doi.org/10.1038/s42256-023-00647-z

جدول داده های توسعه یافته **6** | عملکرد بازیابی پیوند **PT** برای دامنه چکیده ها

روش	λSI	LRAP	nDCG	Ap	MRR
خط پایه تصادفی	N / A	%0.89	%26.0	%0.27	0.016
SciBERT[53]	N / A	%17.22	%52.8	%5.16	0.272
LM PT Baseline (SciBERT مقداردهی اولیه)	0	%16.79	%35.4	%5.00	0.271
DAPT CS RoBERTa[59]	N / A	%32.56	%50.3	%12.86	0.459
LM PT Baseline (CS RoBERTa مقداردهی اولیه)	0	%30.58	%48.3	%12.36	0.438
SIPT (SciBERT مقداردهی اولیه)	0.01	%42.26	%58.7	%14.23	0.536
	0.10	%34.73	%52.5	%9.39	0.457
	0.50	%32.85	%50.8	%8.37	0.438
	0.90	%31.61	%49.8	%7.82	0.426
	0.99	%30.72	%49.0	%6.80	0.415
SIPT (CS RoBERTa مقداردهی اولیه)	0.01	%33.32	%51.2	%8.61	0.448
	0.10	%25.46	%44.4	%5.88	0.359
	0.50	%25.08	%44.0	%6.08	0.355
	0.90	%22.43	%41.6	%4.27	0.317
	0.99	%22.38	%41.5	%4.68	0.316

PT عملکرد بازیابی پیوند را برای یک خط پایه تصادفی، مدل خام SciBERT و SIPT برای پارامترهای مختلف وزن دهی λSI در مجموعه داده مقالات علمی تنظیم کرد. LRAP، میانگین دقت رتبه بندی برچسب. nDCG، سود تجمعی با تخفیف نرمال شده. AP، دقت متوسط؛ MRR، میانگین رتبه متقابل. مقادیر بالاتر نشان دهنده عملکرد بهتر است. تحقق چارچوب SIPT که نتایج بهتری نسبت به قوی ترین خط پایه به همراه دارد، برجسته شده است و شواهدی را ارائه می دهد که نشان می دهد ترکیب اطلاعات رابطه ای در سطح توالی در PT (یعنی $\lambda SI > 0$) منجر به بهبود عملکرد می شود.