



بنام خدا



دانشکده مهندسی برق و کامپیوتر

پردیس دانشکده های فنی

دانشگاه تهران

پیشنهاد و فرم حمایت از پایان نامه تحصیلات تکمیلی

☐ دکتری

☒ کارشناسی ارشد

☐ نوبت دوم

☒ روزانه

نام و نام خانوادگی: محسن فیاض

شماره دانشجویی: ۸۱۰۱۰۰۵۲۴

استاد راهنمای اول: دکتر یدالله یعقوبزاده

تاریخ تصویب:

ورود به تحصیلات تکمیلی پردیس:

شماره مرجع:

۱. شماره‌ی مرجع، توسط معاونت پژوهشی پردیس دانشکده‌های فنی، هنگام صدور ابلاغ، درج خواهد شد.

۲. تکمیل کلیه قسمت های فرم بصورت تایپ شده ضروری می باشد و در صورت ناقص بودن هریک از موارد بدون

هیچ اقدامی پروپوزال ارسال شده عودت داده خواهد شد.

۱- خلاصه‌ی اطلاعات پایان نامه

بررسی دانش و توجه مدل‌های زبانی بافتاری			عنوان (فارسی)
Investigating Knowledge and Attention of Contextual Language Models			عنوان (انگلیسی)
<input type="checkbox"/> توسعه ای	<input checked="" type="checkbox"/> کاربردی	<input checked="" type="checkbox"/> بنیادی	نوع

۲ اطلاعات استاد راهنما و مشاور

نوع مسئولیت	نام و نام خانوادگی	مرتبه‌ی علمی	درصد مشارکت	محل خدمت	امضا و تاریخ
استاد راهنمای اول (مجری)	دکتر یدالله یعقوب‌زاده	استادیار	۱۰۰	دانشگاه تهران	
استاد راهنمای دوم (حسب نیاز)		انتخاب نمایید			
استاد مشاور اول		انتخاب نمایید			
استاد مشاور دوم (حسب نیاز)		انتخاب نمایید			

۳ اطلاعات دانشجو

نام و نام خانوادگی	محسن فیاض	تلفن ثابت/همراه	۰۹۳۵۹۸۴۲۲۵۲
مقطع و نوع پذیرش	کارشناسی ارشد - روزانه	آدرس ایمیل	mohsen.fayyaz ^{۷۷} @ut.ac.ir
رشته و گرایش	مهندسی کامپیوتر هوش مصنوعی و رباتیک	امضا و تاریخ	۱۴۰۱/۰۸/۲۰ 

۴- مشخصات موضوعی پایان نامه

۴-۱ تعریف مساله، هدف و ضرورت انجام (حداکثر سه صفحه)

با توجه به افزایش روزافزون قابلیت‌ها و ظرفیت محاسباتی رایانه‌ها، در کنار تولید داده‌های فراوان در مقیاس بزرگ با کمک همه‌گیری اینترنت و شبکه‌های اجتماعی، استفاده از تکنیک‌های یادگیری ماشین برای مدیریت و تحلیل این حجم از داده ضروری است. در این مسئله، اهمیت متن با وجود داده‌های زبانی بسیار و پیچیدگی‌های تحلیل آن از اهمیت دوچندان برخوردار است.

برای مدیریت و تحلیل این حجم از داده متنی، مدل‌های زبانی متفاوتی با معماری‌های مختلف در طول زمان ارائه و استفاده شده‌اند. اما مدل‌هایی که امروزه بسیار مورد توجه هستند و بهترین نتایج را کسب می‌کنند، مدل‌های زبانی بافتاری و مخصوصاً مدل‌های مبتنی بر معماری ترنسفورمر^۱ هستند. این مدل‌ها توانسته‌اند از زمان معرفی تا کنون، اختلاف چشم‌گیری در نتایج روی مسائل و مجموعه‌داده‌های مختلف داشته باشند. به این ترتیب که بهترین نتایج در محک‌های مختلف کنونی، همگی در اختیار این مدل‌ها قرار گرفته‌اند.

اما این برتری عملکرد بدون خسارت نبوده است. معماری‌های کنونی بر پایه شبکه‌های عصبی عمیق استوار شده‌اند که در معاوضه بین تفسیرپذیری و نتایج بهتر به سمت نتایج بهتر سوق داده شده‌اند. به علت طبیعت مبهم شبکه‌های عصبی، تفسیر عملکرد و رفتار درونی آن‌ها و به طبع آن کاوش مدل‌های زبانی بافتاری امروزی چالش برانگیز است.

تفسیر عملکرد و رفتار مدل می‌تواند باعث آشکار شدن مشکلات موجود، اشکال‌زدایی مدل، آگاهی بخشی در مورد مدل و حتی داده‌ها شود. ضمناً در سناریوهای حساس‌تر که انسان نیز در حلقه تصمیم‌گیری قرار دارد^۲، می‌تواند از علل توضیح داده شده توسط مدل بهره برد. این توضیحات همچنین می‌تواند تعصب و جانبداری مدل‌ها را کشف کند تا در نهایت در مسیر برطرف کردن آن‌ها قدم برداریم. در مجموع، پیشرفت با دید بازتر و آگاهی کامل‌تر در گرو بررسی و تحلیل مدل‌های کنونی است.

بر اساس اهمیت این مبحث، حجم وسیعی از کارهای تحقیقاتی در حوزه تجزیه و تحلیل در پردازش زبان‌های طبیعی انجام شده است. اما با این وجود، هنوز فاقد چارچوب و یا روش‌شناسی مشترک و آزموده شده‌ای که بر سر آن اتفاق نظر باشد هستیم. بنابراین برای رسیدن به هدف نهایی که تجزیه و تحلیل شبکه‌های عصبی مدرن است، لازم است علاوه بر اهداف غایی مثل بررسی دانش‌های زبانی موجود در مدل‌ها و بررسی توجه آن‌ها به اجزای جمله، ابتدائاً بررسی دقیقی روی ابزارهای موجود و در صورت لزوم بهبود و حتی ارائه روش‌های جدید برای آزمایش‌های مورد نیاز داشته باشیم.

در این پروژه، به صورت کلی در راستای تفسیرپذیری و تحلیل مدل‌های زبانی حرکت می‌کنیم. در این مسیر به بررسی دانش‌های زبانی، شامل دانش‌های پایه مانند تشخیص اجزای زبان یا دانش‌های سطح بالاتر مانند وجود استعاره در کنار بررسی توجه به اجزای جمله در مدل‌های زبانی بافتاری مبتنی بر ترنسفورمر می‌پردازیم. در صورت نیاز به طراحی روش‌های مناسب برای این تحلیل‌ها نیز خواهیم پرداخت.

^۱ Transformers

^۲ Human-in-the-loop

۲-۴ روش و فنون اجرایی

در این پژوهش، ابتدا باید خود مدل‌های زبانی نوین مورد بررسی دقیق قرار گیرند و تفاوت‌های جزئی آن‌ها بررسی شود تا تحلیل نتایج نهایی را بتوان به خوبی انجام داد و تحت تاثیر عوامل ناشناخته در مدل از دید نویسندگان نباشد. این بررسی اولیه شامل معماری خود ترنسفورمر و مدل‌های مبتنی بر آن می‌شود تا جزئیاتی مانند ساختار شبکه، اهداف استفاده شده در آموزش آن‌ها و نسبت نتایج آن‌ها به هم دیگر مشخص شود.

سپس روش‌ها و چارچوب‌های پیشین برای تجزیه و تحلیل مدل‌های زبانی مورد مطالعه و بررسی دقیق قرار خواهد گرفت. پس از بررسی به پیاده‌سازی این ابزارها می‌پردازیم و نقاط قوت و ضعف آن‌ها را شناسایی می‌کنیم. در صورتی که یک ابزار مناسب بود، با همان به ادامه تحلیل می‌پردازیم، در غیر این صورت به دنبال روش‌های جایگزین یا اصلاح روش‌های موجود خواهیم رفت. این مرحله نیز اهمیت بالایی دارد زیرا هرگونه جانبداری در معیار بررسی می‌تواند نتایج گمراه‌کننده‌ای به همراه داشته باشد.

در مرحله بعدی، دانش و توجه مدل‌ها را با استفاده از ابزارهای مشخص شده بررسی می‌کنیم. این بخش شامل مطالعه مسائل موجود که می‌تواند دانش مدل را به خوبی محک بزند، مقدار و توزیع دانش‌ها در لایه‌های مختلف و تاثیر آموزش مدل در تغییر مقدار این دانش‌ها است. ضمناً برای توجه مدل می‌توان توجه را در مسائل مختلف و در لایه‌های متفاوت مدل بررسی کرد. چون بخش نهایی این تحقیق به بررسی نتایج به دست آمده و سپس تعیین قدم بعدی بر اساس اطلاعات به دست آمده است، بخش‌هایی از آزمایش‌ها هم از الان مشخص نیستند که در طول پژوهش روشن خواهند شد.

۳-۴ پیشینه پژوهش (همراه با ذکر منابع اساسی)

مدل‌های زبانی بافتاری بعد از معرفی ترنسفورمر [۱] انقلابی در حوزه زبان‌های طبیعی ایجاد کردند. پس از معرفی ترنسفورمر، مدل برت^۳ [۲] با پشت هم قرار دادن چندین لایه از بخش کدگذار^۴ ترنسفورمر توانست نتایجی به دست آورد که اختلاف چشمگیری با نتایج مدل‌های قبلی داشت. در ادامه مدل‌های دیگری با تغییرهایی در معماری مدل، اهداف آموزش و جزئیات دیگر معرفی شدند مانند مدل‌های RoBERTa [۳]، ELECTRA [۴]، ALBERT [۵] و XLNet [۶].

در مورد دانش موجود در بازنمایی‌ها تحقیقاتی با استفاده از کاوندهای^۵ متفاوت انجام شده است. وظیفه این کاوندها مشخص کردن میزان حضور اطلاعاتی خاص درون بازنمایی‌های مدل است.

^۳ BERT

^۴ Encoder

^۵ Probe

از ساده‌ترین کاوندها، کاوند ساختاری [۷] است که با آموزش یک تبدیل خطی روی بازنمایی‌های به دست آمده از برت، نشان می‌دهد که می‌توان فواصل درخت نحوی^۶ را یافت.

با توجه به اینکه لزوماً دانش کدگذاری شده در بازنمایی‌ها به صورت خطی قابل استخراج نیستند، کاوند یال^۷ [۸] معرفی شده است. این کاوند یک شبکه عصبی کم عمق روی بازنمایی‌های ثابت مدل آموزش می‌دهد تا مسائلی مانند تشخیص اجزای کلام را انجام دهد. دقت این طبقه‌بند میزان دانش مورد بررسی در بازنمایی‌ها را نشان می‌دهد. در این مقاله، نشان داده شد که مدل برت بیشتر از مدل‌های بافتاری قبل خود مثل ELMO [۹] و مدل‌های غیر بافتاری دانش زبانی دارد.

در ادامه کاوند یال تحقیقی [۱۰] انجام شد و لایه‌های برت بررسی شدند و میزان دانش در هر لایه بررسی شد. در این تحقیق دیده شد که مسائلی که معنایی تر هستند در لایه‌های بالاتر تجمع دارند و مسائل و دانش‌های نحوی در لایه‌های ابتدایی موجودند.

با استفاده از کاوند یال و ابزارهای تکمیلی، تغییر دانش مدل در طول آموزش نیز بررسی شده است [۱۱]. در این مقاله مشخص شده است که مدل برت محافظه کار است و دانش‌ها تا حد خوبی باقی می‌مانند و ضمناً این صرفاً چند لایه آخر مدل است که دچار تغییرات شدید می‌شود و باقی لایه‌های اولیه تغییر ناچیزی دارند.

در مورد تمام این کاوندها این مسئله وجود دارد که سختی استخراج دانش در نظر گرفته نمی‌شود. به این صورت که با یک شبکه عصبی ثابت سعی داریم تا بازنمایی‌ها را مقایسه کنیم. اما ممکن است استخراج یک دانش راحت‌تر و یکی سخت‌تر باشد که این موضوع به خوبی در این دو کاوند به خوبی نشان داده نمی‌شود. به همین دلیل کاوند MDL معرفی شد [۱۲]. این کاوند بر اساس مفاهیم تئوری اطلاعات سعی دارد علاوه بر دقت نهایی کاوند، سختی رسیدن به آن دقت را نیز در نظر بگیرد. برای این کار دو روش پیشنهاد می‌دهد که یکی استفاده از مدلی با پیچیدگی متغیر است تا میزان پیچیدگی مدل را در میزان دانش تاثیر دهد. روش دوم سعی می‌کند مسئله را به صورت ارسال اطلاعات ببیند و به صورت افزایشی روی داده‌ها آموزش داده و دقت را اندازه‌گیری می‌کند. در روش دوم شهود این است که اگر استخراج دانش از بازنمایی‌ها ساده باشد، باید بتوان با تعداد کمی از نمونه‌های داده الگوی موجود را کشف کرد.

همچنین برای بررسی خود کاوندها نیز تحقیقی [۱۳] انجام شد که با تعریف مسئله کنترل، میزان یادگیری خود کاوند را می‌سنجید و یک کاوند مناسب نباید خودش ظرفیت یادگیری مسئله را داشته باشد، بلکه صرفاً باید دانش بازنمایی را به دست آورد. نشان داده شده است که با این معیار، کاوند MDL بهتر از کاوند یال عمل می‌کند.

در کنار تحقیقات ذکر شده در مورد دانش مدل، در مورد توجه آن به بخش‌های جمله نیز تحقیقات فراوانی شده است. در یکی از اولین تحقیقات [۱۴] که در سال ۲۰۱۹ ارائه شد وزنه‌های مکانیزم توجه مدل بررسی شدند. در این مقاله نشان داده شد که یک سری از سرهای توجه در لایه‌های خاص به نکات به خصوصی مثل علامت‌ها، ضمیرها و مقصودشان، مفعول افعال، و غیره توجه می‌کنند. این تحقیق محدود به تک لایه‌ها بوده و در مورد کل مدل نظری ندارد.

^۶ Syntax Tree

^۷ Edge Probe

در یک کار تکمیلی [۱۵] با استفاده از روشهایی به نام *attention flow* و *attention rollout* جریان توجه مدل در لایه‌ها تجمیع شدند و یک توجه جامع در مورد کل مدل به دست آمد. ضمناً نشان داده شد که این توجه به دست آمده همبستگی بیشتری نسبت به توجه تک لایه‌ها با توجهی که از روش‌های حذف کلمه یا گرادیان به دست می‌آیند دارند و می‌تواند توجه کل مدل را بهتر نمایش دهد.

اما در مقاله دیگری [۱۶] نشان داده شد که وزن‌های توجه مدل به تنهایی نشان‌دهنده توجه واقعی در هر لایه نیست. در این مقاله نشان داده شد که باید به جای وزن توجه، از اندازه بردارهایی که پس از مکانیزم توجه ساخته می‌شود استفاده کرد. به این ترتیب که ممکن است یک بردار خودش بزرگ بوده و وزن توجه کم، صرفاً برای جبران اندازه ابتدایی بزرگ بوده است نه توجه کمتر به آن بردار. وقتی اندازه بردار در نظر گرفته شود، میزان تاثیر واقعی به دست می‌آید که به اندازه ابتدایی بردار وابسته نیست.

البته با توجه به معماری کدگذار ترنسفورمر، به جز بخش مکانیزم توجه، لایه نرمال‌سازی، اتصال باقی‌مانده و یک شبکه عصبی نیز وجود دارد که در نظر گرفته نشده است. در ادامه تحقیق قبلی، مقاله‌ای [۱۷] نوشته شد که لایه نرمال‌سازی و اتصال باقی‌مانده را نیز در نظر گرفت و توانست تخمین دقیق‌تری در هر لایه به دست آورد.

اما همانطور که دیده می‌شود این حوزه تحقیقاتی هنوز هم بسیار نیازمند بررسی و تحلیل بیشتر و ابزارها و تفاسیر دقیق‌تر است که از اهداف این تحقیق است.

- [١] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. u. Kaiser and I. Polosukhin, "Attention is All you Need," in *Advances in Neural Information Processing Systems*, ٢٠١٧.
- [٢] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis, ٢٠١٩.
- [٣] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer and V. Stoyanov, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," *CoRR*, vol. abs/١٩٠٧.١١٦٩٢, ٢٠١٩.
- [٤] K. Clark, M.-T. Luong, Q. V. Le and C. D. Manning, "ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators," in *ICLR*, ٢٠٢٠.
- [٥] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma and R. Soricut, "ALBERT: A Lite BERT for Self-Supervised Learning of Language Representations," in *ICLR 2020*, ٢٠٢٠.
- [٦] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov and Q. V. Le, "XLNet: Generalized autoregressive pretraining for language understanding," in *Advances in neural information processing systems*, ٢٠١٩.
- [٧] J. Hewitt and C. D. Manning, "A Structural Probe for Finding Syntax in Word Representations," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis, ٢٠١٩.
- [٨] I. Tenney, P. Xia, B. Chen, A. Wang, A. Poliak, R. T. McCoy, N. Kim, B. V. Durme, S. R. Bowman, D. Das and E. Pavlick, "What Do You Learn from Context? Probing for Sentence Structure in Contextualized Word Representations," in *International Conference on Learning Representations*, ٢٠١٩.
- [٩] M. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee and L. Zettlemoyer, "Deep Contextualized Word Representations," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, New, ٢٠١٨.
- [١٠] I. Tenney, D. Das and E. Pavlick, "BERT Rediscovered the Classical NLP Pipeline," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, ٢٠١٩.

- [11] A. Merchant, E. Rahimtoroghi, E. Pavlick and I. Tenney, "What Happens To BERT Embeddings During Fine-tuning?," in *Proceedings of the Third BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP*, Online, 2020.
- [12] E. Voita and I. Titov, "Information-Theoretic Probing with Minimum Description Length," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Online, 2020.
- [13] J. Hewitt and P. Liang, "Designing and Interpreting Probes with Control Tasks," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, 2019.
- [14] K. Clark, U. Khandelwal, O. Levy and C. D. Manning, "What Does BERT Look at? An Analysis of BERT's Attention," in *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, Florence, 2019.
- [15] S. Abnar and W. Zuidema, "Quantifying Attention Flow in Transformers," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online, 2020.
- [16] G. Kobayashi, T. Kuribayashi, S. Yokoi and K. Inui, "Attention is Not Only a Weight: Analyzing Transformers with Vector Norms," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Online, 2020.
- [17] G. Kobayashi, T. Kuribayashi, S. Yokoi and K. Inui, "Incorporating Residual and Normalization Layers into Analysis of Masked Language Models," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Online and Punta Cana, Dominican Republic, 2021.

شماره:		
تاریخ:		
<p>معاون محترم آموزشی و تحصیلات تکمیلی پردیس دانشکده‌های فنی</p> <p>با سلام و احترام،</p> <p>فرم پیشنهاد و حمایت از پایان نامه جناب آقای : محسن فیاض با عنوان: بررسی دانش و توجه مدل‌های زبانی بافتاری به راهنمایی جناب آقای دکتر: یدالله یعقوب‌زاده که در راستای برنامه جامع تحقیقات ایشان با عنوان: هوش ماشین و رباتیک میباشد</p> <p>نمیباشد در تاریخ در شورای پژوهشی و تحصیلات تکمیلی دانشکده مهندسی برق و کامپیوتر به تصویب رسید. خواهشمند است، دستور فرمایند اقدام لازم انجام شود.</p> <p>معاون پژوهشی و تحصیلات تکمیلی دانشکده مهندسی برق و کامپیوتر</p> <p>امضا:</p> <p>تاریخ</p>		

شماره:		
تاریخ:		
<p>معاون محترم پژوهشی پردیس دانشکده‌های فنی</p> <p>با سلام و احترام،</p> <p>به پیوست، فرم پیشنهاد و حمایت از پایان نامه تحصیلات تکمیلی به همراه مشخصات آن، که به تصویب شورای پژوهشی تحصیلات تکمیلی دانشکده مهندسی برق و کامپیوتر رسیده است جهت دستور به اقدام مقتضی تقدیم می‌شود.</p> <p>معاون آموزشی و تحصیلات تکمیلی پردیس دانشکده‌های فنی</p> <p>امضا:</p> <p>تاریخ</p>		