دانشکده مهندسی کامپیوتر

دستیار آزمایشگاه

آزمایشگاه پردازش زبان طبیعی دانشگاه علم و صنعت ایران

غزاله محمودی

نام استاد کارآموزی:

دکترمحمد طاهر پیله‌ور

اردیبهشت ماه ۱۴۰۰


بسم الله الرحمن الرحیم

تأییدیه‌ی صحت و اصالت نتایج

بسمه تعالی

اینجانب غزاله محمودی به شماره دانشجویی ۹۶۵۲۲۲۴۹ دانشجوی رشته مهندسی کامپیوتر مقطع تحصیلی کارشناسی تأیید می‌نمایم که کلیه‌ی مطالب مندرج در این گزارش حاصل ۳۰۰ ساعت حضور و کار اینجانب در شرکت/کارخانه آزمایشگاه پردازش زبان طبیعی دانشگاه علم و صنعت ایران و بدون هرگونه دخل و تصرف است و موارد نسخه‌برداری‌شده از آثار دیگران را با ذکر کامل مشخصات منبع ذکر کرده‌ام. درصورت اثبات خلاف مندرجات فوق، به تشخیص دانشگاه مطابق با ضوابط و مقررات حاکم آموزشی، پژوهشی و انضباطی با اینجانب رفتار خواهد شد و حق هرگونه اعتراض درخصوص احقاق حقوق مکتسب و تشخیص و تعیین تخلف و مجازات را از خویش سلب می‌نمایم.

نام و نام خانوادگی : غزاله محودی

امضا و تاریخ : ۱۴۰۰/۰۲/۰۹

تشکر و قدردانی:

با تشکر از استاد سید صالح اعتمادی و خانم مریم سادات هاشمی که اینجانب مراحل کارآموزی خود را تحت سرپرستی ایشان گذراندم.

چکیده

در سال‌های اخیر و با پیشرفت‌های چشمگیر در حوزه هوش مصنوعی، پردازش زبان طبیعی و پردازش تصویر مسئله‌هایی با کاربرد عملی در زندگی روزمره انسان‌ها طراحی شده است. یکی از مواردی که اخیرا مورد توجه قرار گرفته است بحث پرسش و پاسخ تصویری می‌باشد. این مسئله کاربردهای زیادی در کمک به نابینایان، دستیار هوشمند و موارد مشابه می‌تواند داشته باشد. این دوره شامل آشنایی و کسب تجربه مرتبط با مسئله پرسش و پاسخ تصویری بود.

واژه‌های کلیدی: پرسش و پاسخ تصویری، VQA، Visual Question Answering.

فهرست مطالب

فهرست اشکال

فصل ۱

معرفی حوزه کارآموزی

۱-۱ مقدمه

آزمایشگاه پردازش زبان طبیعی به سرپرستی دکتر اعتمادی مشغول به انجام پروژه‌ها با محوریت هوش مصنوعی و پردازش زبان طبیعی می‌باشد. با توجه به شرایط همه‌گیری ویروس کرونا و غیر حضوری شدن دانشگاه‌ها برگزاری جلسات و هم‌اندیشی‌ها به صورت کاملا مجازی در بستر Microsoft Teams صورت می‌گیرد.

در آزمایشگاه دانشجویان تحصیلات تکمیلی مشغول کار بر روی پایان‌نامه می‌باشند و تعدادی کارآموز به عنوان دستیار به آن‌ها در انجام امور محوله کمک می‌کنند و دانش و تجربه عملی کسب می‌کنند.

۱-۱-۱ پرسش و پاسخ تصویری

در سال‌های اخیر و با پیشرفت‌های چشمگیر در حوزه هوش مصنوعی، پردازش زبان طبیعی و پردازش تصویر مسئله‌های با کاربرد عملی در زندگی روزمره انسان‌ها طراحی شده است. یکی از مواردی که اخیرا مورد توجه قرار گرفته است بحث پرسش و پاسخ تصویری می‌باشد. در این مسئله، یک تصویر و سوالی مربوط به تصویر به عنوان ورودی به مدل داده می‌شود و سیستم با تحلیل تصویر و پردازش متن ورودی پاسخ مناسبی به سوال مطرح شده می‌دهد.

فصل ۲

مشروح فعالیت‌های انجام شده در محل استقرار

۲-۱ مقدمه

فعالیت‌های صورت گرفته در مدت کارآموزی بر محوریت مسئله پرسش و پاسخ تصویری می‌باشد. در ابتدا آموزش‌های لازم مرتبط با هوش مصنوعی و یادگیری ماشین فرا گرفته شد. در ادامه ارزیابی دیتاست ترجمه شده VQA v1 انجام شد. سپس راه‌کارهای جمع آوری دیتاست فارسی مورد بررسی قرار گرفت. همچنین وب‌سایت پرسش و پاسخ تصویری برای دسترسی عمومی روی سرور راه اندازی شد و در نهایت با توجه به اطلاعات کسب شده و تسلط بیشتر به کلیت مسئله مقاله‌ای با محوریت هرس شبکه LXMERT [1] تاثیر آن بر نتایج مسئله VQA [۲] نوشته شد.

۲-۲ شرح پروژه‌ها و فعالیت های انجام‌شده توسط کارآموز

فعالیت‌های صورت گرفته در این دوره به چندین بخش تقسیم می‌شود که در ادامه به شرح مختصری از هر کدام از موارد پرداخته می‌شود.

۲-۲-۱ گذراندن دوره آموزشی Coursera Deep Learning Specialization

در ابتدا برای آشنایی بیشتر با مباحث یادگیری عمیق و شبکه های عصبی تصمیم بر آن شد که دوره آموزشی معتبر در این باره گذرانده شود. با تحقیقات انجام شده دوره آموزشی Deep Learning Specialization برای deeplearning.ai با تدریس Andrew Ng انتخاب شد و ویدیوها و تمارین و کوئیزهای مربوط به آن انجام شد که محتوا آن شامل ۵ زیر بخش می‌باشد.

- Neural Networks and Deep Learning
- Improving Deep Neural Networks: Hyperparameter Tuning, Regularization and Optimization
- Structuring Machine Learning Project
- Convolutional Neural Networks
- Sequence Models

۲-۲-۲ ارزیابی ترجمه ماشینی

همان طور که قبلا ذکر شد در این دوره مسائلی حول تسک VQA [۲] انجام شده است. دیتاست VQA v1 به زبان انگلیسی می‌باشد. از این رو برای انجام تسک برای زبان فارسی می‌بایست یا دیتاست فارسی جمع‌آوری کرد و یا دیتاست موجود را ترجمه کرد. در این بخش دیتاست ترجمه شده به کمک دو ترجمه‌گر ماشینی Google و ترگمان موجود بود. وظیفه ما ارزیابی صحت ترجمه و شباهت آن با ترجمه انسانی بود. در ادامه معیارهای ارزیابی ترجمه معرفی می‌شود و نتایج آن بر روی دیتاست ترجمه شده بررسی می‌شود.

۱-۲-۲-۲- معرفی دیتاست VQA v1

مجموعه داده‌ای که برای تست انتخاب شده بود، دیتاست VQA v1 است. در این مجموعه برای هر تصویر سه سوال وجود دارد. نوع اول سوال بله و خیر است. نوع دوم مربوط به تعداد شی در تصویر است و نوع سوم شامل سوالات دیگر است. به ازای هر سوال ۱۰ پاسخ وجود دارد. برای ترجمه مجموعه داده از دو ابزار Google [1] و ترگمان[2] استفاده شده است.

۲-۲-۲-۲- معرفی معیار های ارزیابی ترجمه ماشینی

۱. معیار BLEU

معیار BLEU [۳] یکی از معروف‌ترین معیارهای ارزیابی خودکار کیفیت متن ترجمه انسانی در مقایسه با ترجمه ماشینی همان متن می‌باشد. این معیار مقداری بین صفر و یک می‌گیرد. هر چه مقدار به یک نزدیک‌تر باشد، کیفیت ترجمه ماشینی و شباهت آن به ترجمه انسانی بیشتر است. این معیار برای با مقایسه میانگین تعدادn-gram های مشابه در ترجمه ماشینی و انسانی میزان شباهت این دو ترجمه را بررسی می‌کند.

$$BLEU: N = BP\ (\prod_{n=1}^{N} \frac{n\text{-gram}(T \cap R)}{n\text{-gram}(T)})^{\frac{1}{N}}$$

فرمول ۱ :BLEU

از مزایا این معیار سادگی در محاسبه آن می‌باشد. همچنین کار با آن بسیار راحت و آسان است.

[1] www.google.com

[2] targoman.ir

البته این معیار مشکلاتی هم دارد از جمله اینکه معانی کلمات را در نظر نمی‌گیرد، به ساختار جمله در زبان‌های مختلف توجهی ندارد و نتیجه ارزیابی آن در مقایسه با ارزیابی انسانی در برخی متون کاملا متفاوت است.

به طور کلی به دست‌آوردن مقدار BLEU [۳] بین ۰.۶ تا ۰.۷ برای یک ترجمه نشان دهنده این است که ترجمه قابل قبول است و احتمالا از واژگان و اصطلاحات متفاوت با معنی یکسان استفاده شده است.

۲. معیار NIST

معیار NIST [۴] نیز یکی از معیارهای ترجمه ماشینی می‌باشد. نحوه محاسبه این معیار شبیه BLEU [۳] می‌باشد و تنها تفاوت در این است که برای n-gram های مختلف وزن‌های متفاوتی می‌توان در نظر گرفت.

۳. معیار METEOR

معیار METEOR [۵] یکی دیگر از معیارهای خودکار ارزیابی کیفیت متن ترجمه شده است. این معیار بر اساس میانگین هارمونیک Recall , Precision محاسبه می‌شود. این معیار علاوه بر رفع مشکلات BLEU [۳] از جمله در نظر گرفتن کلمات مترادف یا هم معنی، تناسب بیشتری با قضاوت‌های انسانی دارد.

$$F_{mean} = \frac{10PR}{R + 9P}$$

$$p = 0.5(\frac{c}{U_m})^3$$

$$M = F_{mean}(1 - p)$$

فرمول ۲: *METEOR*

۴. معیار GTM

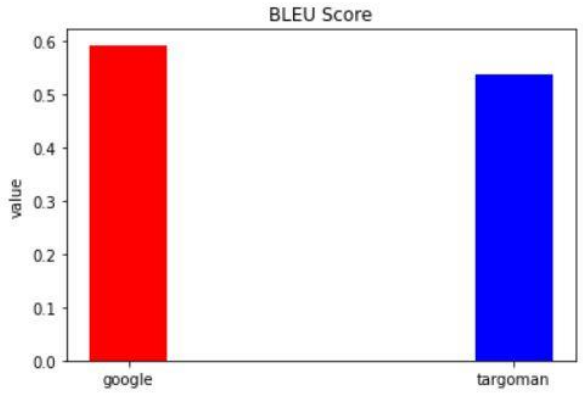
معیار GTM یکی دیگر از معیارهای ارزیابی خودکار کیفیت متن ترجمه شده است. این معیار با محاسبه Recall, Precision و F-measure در هنگام بیشینه تشابه n-gram ها محاسبه می‌شود.

۳-۲-۲-۲- نتایج و تحلیل

۱. معیار BLEU

با مقایسه دو ترجمه ترگمان و Google با معیار BLEU [3] مشاهده شد این دو ترجمه امتیاز نزدیک به هم می‌گیرند و البته امتیاز ترجمه Google کمی بیشتر است. همچنین

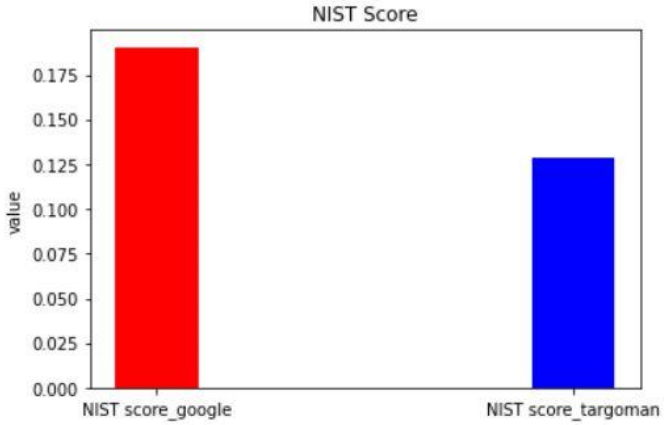
مقدار نزدیک ۰.۶ به دست آمد که نشان دهنده قابل قبول بودن ترجمه می‌باشد.



تصویر ۱ : نتایج معیار BLEU بر روی دیتاهای ترجمه شده

۲. معیار NIST

نتایج این قسمت همچون قسمت قبل به دست آمد و Google امتیاز بیشتری کسب کرد.



تصویر ۲ : نتایج معیار NIST بر روی دیتاهای ترجمه شده

۳-۲-۲ راه‌اندازی وب‌سایت برای پرسش و پاسخ تصویری

طی جلسات گروهی وب‌سایت از قبل موجود برای این مسئله توسط اعضا ارتقا پیدا کرد و با همکاری اعضاء گروه بر روی سرور آزمایشگاه راه‌اندازی شد تا بر بستر اینترنت در دسترس باشد.[1] در پیاده‌سازی قسمت فرانت‌اند از html, css, java script و برای قسمت سرور از flask استفاده شده است.

[1] http://194.225.229.203

تصویر ۳ : وب‌سایت پرسش و پاسخ تصویری

۲-۲-۴ طراحی بات تلگرام برای جمع‌آوری دیتاست فارسی

همان‌طور که اشاره شد برای پرسش و پاسخ تصویری در زبان فارسی دیتاست موجود نمی‌باشد. از این رو برای یافتن روش مناسب جمع‌آوری دیتاست، بحث و بررسی صورت گرفت. راه‌های مختلفی از جمله درست کردن وب‌سایت، بازی تعاملی و بات تلگرام گزینه‌های پیشنهادی بود . با بررسی‌های انجام شده تصمیم بر آن شـــد که بات تلگرام برای جمع‌آوری دیتا راه اندازی شـــود چون زمان کمتری برای انجام آن نیاز بود و در دسترس‌تر از سایر موارد پیشنهادی به نظر می‌رسید. برای راه اندازی بات سه سناریو مد نظر بود که در اینجا به توضیح سناریو پیاده‌سازی شده می‌پردازم.

در این سـناریو با شـروع کار از جانب کاربر یک تصـویر همراه با سـوال برای کاربر ارسـال می‌شـود و کاربر تنها باید به سـوال پاسـخ دهد. همچنین کار دیگر طراحی دیتابیس برای جمع‌آوری بهینه دیتا و ذخیره در سـرور بود. پس از طراحی دیتابیس اولیه توسط اینجانب، توسط بقیه اعضاء گروه مورد بررسی قرار گرفت. کد اولیه بات تلگرام با زبان پایتون نوشـته شـد. کد این پیاده‌سـازی گیت‌لب پروژه موجود اسـت.[1] در ادامه تلاش کردیم

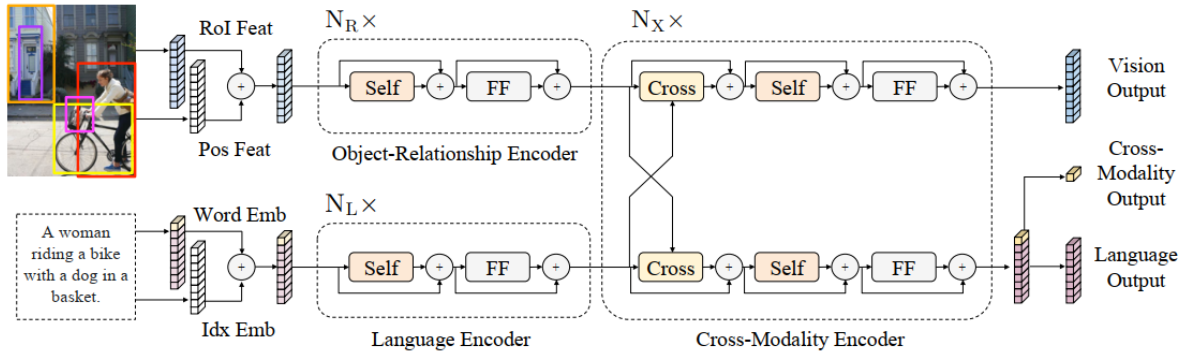
[1] https://gitlab.com/maryamhashemi/pvqa-dataset

نوع پیاده‌سازی را با فریمورک جنگو تغییر دهیم. در این مسیر موارد مختلفی توسط همه اعضا بررسی شد.

تصویر ۴ : بات تلگرام پرسش و پاسخ تصویری

۲-۲-۵ مقاله LXMERT Model Compression for Visual Question Answering

در پخش پایانی دوره کارآموزی مشغول آماده‌سازی مقاله با موضوع LXMERT Model Compression for Visual Question Answering شـدیم. در این مقاله تاثیر هرس شـبکه LXMERT [1] بر دقت مدل بر مسئله VQA و مقایسه آن با شبکه کامل LXMERT [1] مورد بررسی قرار گرفت.



تصویر ۵: مدل LXMERT[۱]

شبکه LXMERT [۱] یک مدل cross-modality و transformer-base می‌باشد که تصویر و متن را به عنوان ورودی می‌گیرد و سه خروجی Vision, Language, Cross-Modality می‌دهد. این شبکه به صورت Pretrain می‌باشد که بر روی چندین دیتاست از جمله QA آموزش دیده است.

با توجه به نظریه Lottery Ticket [6] تصمیم برآن شد که شبکه LXMERT [۱] را به روش‌های مختلف هرس کرده و نتایج به دست‌آمده ارزیابی شود. در ادامه به توضیح مختصری از فرآیند هرس و زیر شبکه‌های تعریف شده پرداخته می‌شود.

فرآیند هرس به این صورت است که به صورت متوالی در هر iteration ، ۱۰ درصد از کوچک‌ترین وزن‌های شبکه دور ریخته می‌شوند. این فرآیند مادامی که x درصد از وزن‌های شبکه باقی بمانند ادامه می‌یابد. شبکه عصبی با pytorch پیاده‌سازی شده است. در قسمتی از پیاده‌سازی و نحوه مکانیزم هرس کردن ابهاماتی به وجود آمد که با ارتباط از طریق ایمیل با Hao Tan و Sai Prasanna، ابهامات به وجود آمده برطرف شد.

وزن‌های باقی‌مانده از هرس Good Subnetwork نامیده شدند. تعدادی از وزن‌هایی از شبکه که دور ریخته شده‌اند به عنوان Bad subnetwork انتخاب شد. همچنین شبکه random نیز تولید می‌شد. لازم به ذکر است سایز این سه شبکه با یکدیگر برابر می‌باشد. در نهایت تاثیر نوع شبکه بر وزن‌ها و دقت مدل در هر کدام از این سه شبکه بررسی شد. [۷]

۳-۲ نتیجه‌گیری

تجربه گذراندن کارآموزی در محیط آکادمیک و انجام کلیه مراحل نوشتن مقاله و بحث و تبادل نظر درباره آن با اعضاء گروه بسیار خوب و مفید بود.

فصل ۳

نتیجه‌گیری و پیشنهادها

۳-۱ مقدمه

تجربه یادگیری و انجام کارگروهی از بزرگ‌ترین مهارت‌هایی اســـت که هر فردی باید آن را فراگیرد. در این دوره این فرصـــت ایجاد شـــد تا علاوه بر تجربه‌های علمی در زمینه مهارت‌های ارتباطی در حل مســـئله نیز تجربیاتی کسب شود. گذراندن این دوره به صورت گروهی همراه با دانشجو ارشد تجربه بسیار خوبی بود. در این دوره توانستیم از همدیگر علاوه بر موارد علمی، نکات اخلاقی مورد نیاز در کار گروهی رو هم یاد بگیریم.

۳-۲ اعلام پیشنهادهایی برای رفع چالش های حوزه/واحد کارآموزی

کار در حوزه هوش مصـنوعی علاوه بر دانش نیاز به سـخت‌افزارهای مناسـب برای اجرا کدها دارد. افرادی که می‌خواهند در زمینه هوش مصنوعی کار کنند باید همه چالش‌های آن از جمله دسترسی به سخت افزارهای مناســـب و زمان طولانی اجرا کدها تا رســـیدن به نتیجه را در نظر بگیرند و صـــبر خود در این موارد افزایش دهند.

فصل ۴

مراجع

مراجع

[١] H. Tan, and M. Bansal, "LXMERT: Learning Cross-Modality Encoder Representations from Transformers," *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP).* pp. 5100-5111.

[٢] A. Agrawal, J. Lu, S. Antol, M. Mitchell, C. L. Zitnick, D. Parikh, and D. Batra, "VQA: Visual Question Answering," *International Journal of Computer Vision,* vol. 123, no. 1 Int. J. Comput. Vision, pp. 4–31,2017.

[٣] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: a Method for Automatic Evaluation of Machine Translation," *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics.* pp. 311-318.

[٤] M. Przybocki, K. Peterson, S. Bronsart, and G. Sanders, "The NIST 2008 Metrics for machine translation challenge—overview, methodology, metrics, and results," *Machine Translation,* vol. 23, no. 2, pp. 71-103, 2009/09/01, 2009.

[٥] S. Banerjee, and A. Lavie, "METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments," *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization.* pp. 65-72.

[٦] J. Frankle, and M. Carbin, "The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks," 2019.

[٧] S. Prasanna, A. Rogers, and A. Rumshisky, "When BERT Plays the Lottery, All Tickets Are Winning," *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP).* pp. 3208-3229.