یادگیری با نمونه های محدود به کمک روش‌های مبتنی بر پیام‌واره

*** مرتضی بهرامی، محرم منصوری زاده، حسن ختن لو**

گروه آموزشی مهندسی کامپیوتر، دانشکده مهندسی دانشگاه بوعلی سینا، همدان

mbahrami520@gmail.com

ستاد ملی هفته پژوهش و فناوری

دانشگاه بوعلی سینا

چکیده

در پردازش زبان طبیعی، امروزه داده های برچسب‌دار اهمیت بالایی دارند و گاهی تعداد در دسترس کم است و گاهی نیز نایاب هستند. بسیاری از کارها وجود دارند که به سختی میتوان داده های مورد نیاز برچسب دار را برای آن تهیه کرد. برای مثال در ترجمه ماشینی، باید داده‌های زیادی در زبان مقصد تهیه کنیم تا عملکرد کار در حد قابل قبولی باشد. ممکن است در زبان مقصد نتوانیم داده‌های مفیدی را جمع آوری کنیم. از این رو نیاز است که از یادگیری با داده‌های کم یا بدون داده استفاده کنیم. اخیرا متدی با عنوان پیام‌واره نویسی (Prompting) معرفی شده که دراین کار، ورودی‌های متنی با استفاده از یک قالب معین، به متن با ساختار جدید تبدیل میشوند که یک جای خالی دارد. در نهایت مدل آموزش دیده، بهترین کلمه را جایگزین جای خالی می کند. این کار، می تواند به ما در زمینه یادگیری با نمونه‌های کم و حتی بدون نمونه کمک کند. پیش‌تر با استفاده از مدل‌های زبانی بزرگی مثل GPT-2 و GPT-3، بدون داده برچسب‌دار و با استفاده از متد پیام‌واره، وظایفی مانند ترجمه را انجام می دادند. اما استفاده از این نوع مدل ها، به علت اینکه تعداد زیادی پارامتر دارند، نیازمند سخت افزارهای بزرگی هستند. در ادامه، PET و IPET معرفی شدند که با استفاده از پیام‌واره نویسی و مدل های کوچکتری مثل Bert و Roberta یادگیری با داده های کم و یا بدون استفاده از داده‌ها را انجام داده اند. به طور مثال در دیتاست Yahoo که یک دیتاست طبقه بندی متن است، با استفاده از IPET و Roberta و ده دیتای برچسب‌دار، به دقت ۷۰ درصدی رسیده است. در این پژوهش میخواهیم با استفاده از پیام‌واره نویسی، به ساختاری دست پیدا کنیم تا بتوان با دیتای کم، بسیاری از وظایف NLP را با دقت بالا و قابل قبولی انجام دهیم.

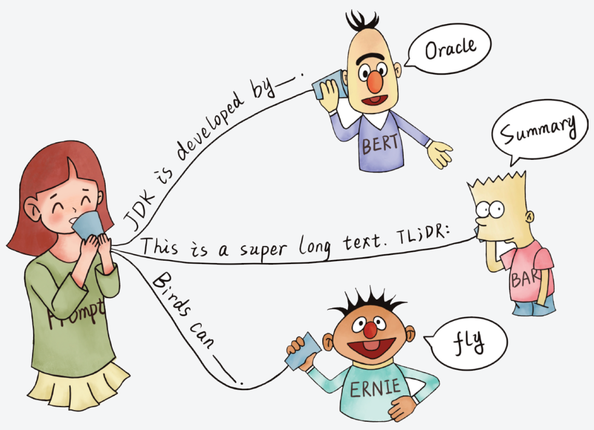
واژه های کلیدی: پردازش زبان طبیعی، یادگیری با نمونه های محدود، پیاو واره، مدل زبانی، Prompting

مقدمه

در سال های اخیر مدل های پیش آموزش دیده زیادی برای انجام وظیفه های مختلفی مانند ترجمه وتحلیل احساسات ساخته شده اند. هر کدام از مدل ها ویژگی های مختص خود را دارد. کارایی مدل ها به مرور بهتر شده و بزرگتر شده اند. اخیرا Transformer ها معرفی شدند که از مفهوم Attention استفاده می کند.

مبدل منجر به پیشرفت قابل توجهی در بسیاری از وظایف چالش برانگیز NLP شده است. مدل های پیش آموزش دیده با تعداد پارامتر های زیاد و بسیار قوی مانند Bert و GPT معرفی شدند. اما هنوز محدودیت هایی وجود دارد، و آن هم این است که هنوز نیاز به مجموعه داده های خاص وظیفه و تنظیم دقیق آن وجود دارد. برای آن که به عملکرد مطلوبی در وظیفه برسیم نیاز است که داده های برچسب دار زیادی در همان وظیفه را آماده سازی کنیم و مدل را تنظیم دقیق کنیم. در واقع سیستم‌های کنونی بیشتر به عنوان متخصصان محدود شناخته می‌شوند تا کلی‌گرایان شایسته. ما میخواهیم بدون نیاز به داده های زیاد و همچنین مدل های بزرگی مانند GPT-3، مدلی واحد طراحی کنیم که بتواند با دیتای کم و برای چند وظیفه مختلف از NLP به دقت مطلوبی برسد.

امروزه برای حذف محدودیت ها، پژوهش های فراوانی انجام شده است. یکی از رویکرد های جدید که به این کار کمک میکند، پیام‌واره است. این متد به سه مرحله ۱) پیش آموزش ۲) پیام‌واره ۳) پیش بینی تقسیم میشود. در این کار ساختار ورودی متنی با استفاده از تابع پیام واره به یک ساختار مشخص تبدیل شده و با استفاده از مدل های از پیش آموزش دیده، پیش بینی و برچسب گذاری انجام می شود.



تصویر۱: تحول در پردازش زبان طبیعی با معرفی Prompting

#	سال	عنوان مقاله	خلاصه کار	خلاصه نتایج
۱	۲۰۱۹	Language models are unsupervised multitask learners	معرفی GPT-2 با ۱/۵ میلیارد پارامتر و انجام یادگیری بدون نمونه	کسب نتایج نسبتا خوب مثلا ۶۰ درصد دقت در دیتاست LAMBADA
۲	۲۰۲۰	Exploiting Cloze Questions for Few Shot Text Classification and NLI	معرفی PET و IPET با استفاده از Roberta و انجام یادگیری با نمونه های محدود	کسب نتایج خوب مثلا ۷۰ درصد دقت در دیتاست Yahoo با ۱۰ دیتا اولیه
۳	۲۰۲۰	Language Models are Few-Shot Learners	معرفی GPT-3 با ۱۷۵ میلیارد پارامتر و انجام یادگیری بدون نمونه با استفاده از Prompting	در بسیاری از وظایف و دیتاست ها به دقت های خوبی رسیده است.
۴	۲۰۲۰	Automatically Identifying Words That Can Serve as Labels for Few-Shot Text Classification	اضافه کردن تعریف اتوماتیک نگاشت ها در PET و معرفی آن با عنوان PETAL	در وظایف مختلف دقت های خوبی کسب کرده است. اما نیاز به بهبود دارد.
۵	۲۰۲۱	Making Pre-trained Language Models Better Few-shot Learners	تنظیم دقیق مدل‌های زبانی کوچک تر با نمونه های محدود و ساخت خودکار قالب ها	به طور متوسط ۱۱ درصد بهبود دقت نسب به GPT-3

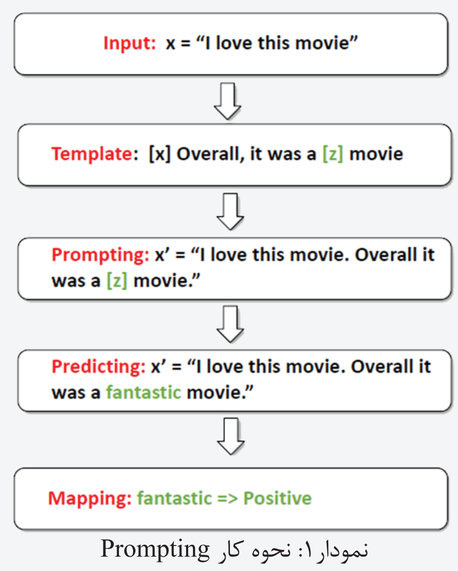
جدول۱: مروری بر کارهای انجام شده به صورت خلاصه

برخی از کارهای انجام شده ضعف هایی دارند. بزرگ بودن مدل زبانی استفاده شده، نیاز به ساخت دستی قالب ها و نگاشت ها و پایین بودن دقت ها از ضعف های عمده کارهای پیشین بوده است. در این پژوهش، هدف بهبود و حل مشکلات این حوزه است.

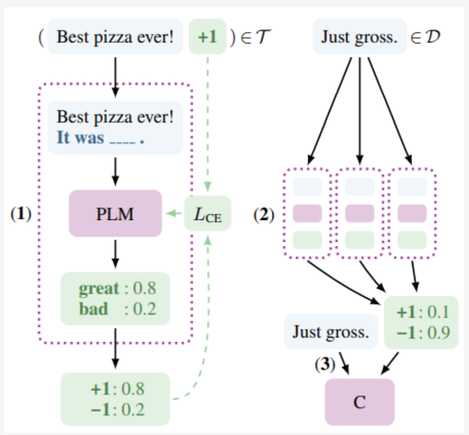
مواد و روش ها

همانطور که گفته شد، هسته پژوهش، متد Prompting است. مطابق تصویر مقابل، یک قالب طراحی می شود و بر اساس آن، ساختار ورودی متنی به یک ساختار مشخص تبدیل میشود که جای خالی در آن وجود دارد. با استفاده از مدل های از پیش آموزش دیده، پیش بینی و برچسب گذاری انجام می شود. نکته مهم کار این است که به داده های زیاد احتیاج نداریم و یا ممکن است نیازی به داده نباشد. از لحاظ انتخاب مدل زبانی پیش آموزش دیده، مهندسی ساخت قالب و مهندسی ساخت نگاشت ها میتوان تاثیر زیادی در خروجی نهایی داشت. این متد، کاربرد های متنوعی در وظایف مختلف NLP دارد و بر اساس هر وظیفه ای که میخواهیم انجام دهیم باید قالب و تابع نگاشت مختص آن را بسازیم.

یکی از ایده ها این است که PET را گسترش دهیم. مطابق تصویر مقابل، در PET، با استفاده از داده های برچسب دار و تغییر آنها بر اساس قالب ساخته شده، تعدادی مدل پیش آموزش دیده به صورت مجموعه ای، تنظیم دقیق می شوند و در نهایت مدل ها می توانند داده های بدون برچسب را با دقت مناسبی به داده برچسب دار تبدیل کنند. داده های تولیدی به مدل یادگیری ماشین تحویل داده می شوند.



نمودار۱: نحوه کار Prompting



تصویر۲: نحوه کار PET

نتایج

مقاله PET به تنهایی به روی بسیاری از دیتاست ها آزمایشات مختلفی انجام داده است. به طور نمونه، دقت دو دیتاست Yelp و AG news در جدول مقابل مقایسه شده است.

|T| تعداد دیتای مورد استفاده است. دقت مدل PET و IPET با رویکرد قدیمی (supervised learning) مقایسه شده است.

Num	Method	Yelp	AG's
|T| = 10	supervised	21.1 ± 1.6	25.0 ± 0.1
	PET	52.9 ± 0.1	87.5 ± 0.0
	IPET	**57.6 ± 0.0**	**89.3 ± 0.1**
|T| = 50	supervised	44.8 ± 2.7	82.1 ± 2.5
	PET	60.0 ± 0.1	86.3 ± 0.0
	IPET	**60.7 ± 0.1**	**88.4 ± 0.1**

جدول۲: مروری بر نتایج PET

منابع

1. Alec Radford et al. Language Models are Unsupervised Multitask Learners. *OpenAI blog*
2. Brown et al. Language Models are Few-Shot Learners. *NeurIPS 2020*, arXiv:2005.14165v4
3. Timo Schick and Hinrich Schütze. Exploiting Cloze Questions for Few Shot Text Classification and Natural Language Inference. *EACL2021*, arXiv:2001.07676v3
4. Gao et al. Making Pre-trained Language Models Better Few-shot Learners. *ACL 2021*, arXiv:2012.15723
5. Timo schick et al. Automatically Identifying Words That Can Serve as Labels for Few-Shot Text Classification. *COLING 2020*, arXiv:2010.13641v1
6. Pengfei Liu et al. Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. arXiv:2107.13586v1

اختصارات

NLP: Natural Language Processing, **PET**: Pattern Exploiting Training
IPET: Iterative Pattern Exploiting Training, **PETAL**: PET with Automatic Labels