

Wang:2020

e:2019

chen:2019

goo-et al-2018-slot

aligned_lstm_atten_nlu

Wang:18

Daha:2019

Siddhant:2019

huang:2020

Qin:2021

Yang:2021

priorknowledge:2023

در معماری $\text{cite}\{\text{chen:2019}\}$ از $\text{Ir}\{\text{BERT}\}$ به عنوان رمزنگار استفاده شد. اما $\text{Ir}\{\text{BERT}\}$ در درک روابط منطقی میان برجسب‌های هدف، خوب عمل نمی‌کند و این مشکل باعث می‌شود عملکرد تشخیص جای خالی شدیداً تحت تأثیر قرار گیرد $\text{cite}\{\text{Wang:2020}\}$. از این رو، مدل $\text{Ir}\{\text{SASGBC}\} \setminus \text{cite}\{\text{Wang:2020}\}$ معرفی شد، که حاوی دو راهکار برای مقابله با مشکل ذکر شده بود. این دو راهکار مکانیزم $\text{Ir}\{\text{Slot-Gate}\}$ و رمزگشای $\text{Ir}\{\text{CRF}\}$ بودند که هر دو مختص رمزگشای تشخیص جای خالی به کار گرفته شدند. در رمزگشای تشخیص جای خالی، نخست مکانیزم $\text{Ir}\{\text{Slot-Gate}\}$ برای آمیختن هدف پیش بینی شده، با خروجی $\text{Ir}\{\text{BERT}\}$ پیشنهاد شد. در مکانیزم $\text{Ir}\{\text{Slot-Gate}\}$ پیشنهادی، خروجی تعبیه‌ی $\text{Ir}\{\text{BERT}\}$ برای هر واژه، با بردار خروجی نشانه‌ی $\text{\$[CLS]\$}$ الحاق شده، و سپس با استفاده از یک لایه‌ی خطی بدون بایاس، ابعاد آن کاهش می‌یابد. دوم، در آخرین لایه از رمزگشا، از $\text{Ir}\{\text{CRF}\}$ برای پیش‌بینی احتمالات برجسب‌ها و برای تولید برجسب‌های خروجی از الگوریتم $\text{Ir}\{\text{Viterbi}\}$ استفاده شد. برای رمزگشای تشخیص هدف، صرفاً یک لایه‌ی خطی همراه با سافت‌مکس، احتمالات خروجی را تعیین می‌کرد.

در روش $\text{Ir}\{\text{Slot-Gate}\}$ ، اطلاعات تگ‌های تولید شده در تشخیص هدف مورد بهره‌وری واقع نمی‌شد؛ از این رو، ارتباطات مستقیم دوطرفه برقرار نمی‌گردید. مدل $\text{Ir}\{\text{Inter-Related}\}$ شبکه‌ی $\text{Ir}\{\text{SF-ID}\}$ را معرفی کرد که از یک زیرشبکه‌ی تشخیص جای خالی و یک زیرشبکه‌ی تشخیص هدف ساخته می‌شد. زیرشبکه‌ی تشخیص جای خالی، اطلاعات تشخیص هدف را وارد تشخیص جای خالی کرده و در مقابل، زیرشبکه‌ی تشخیص هدف، اطلاعات برجسب‌های تولید شده را بر شبکه‌ی تشخیص هدف اعمال می‌کرد. اما این معماری به مثابه قضیه‌ی مرغ و تخم مرغ است! برای تولید هر کدام، نیاز به داشتن دیگری است. از این رو، این مدل، مکانیزم تکرار را معرفی کرد که در آن، یک وظیفه به عنوان نقطه‌ی آغاز انتخاب می‌شد، اما فرایند انتخاب تا تعداد دفعه‌ی مشخصی تکرار می‌شد. به این ترتیب اطلاعات معنایی هر دو وظیفه می‌توانستند با یکدیگر ترکیب شوند. از نظر ساختار شبکه‌ای، این مدل در رمزنگار خود یک $\text{Ir}\{\text{LSTM}\}$ دوطرفه با مکانیزم توجه جداگانه برای هر وظیفه تعریف کرد. در رمزگشای آن، مکانیزم تکرار، درواقع لایه‌های تماماً متصل، و برای تولید برجسب از $\text{Ir}\{\text{CRF}\}$ استفاده کرد.

طراحی مدل زبانی $\text{Ir}\{\text{BERT}\}$ به نحوی است که استفاده از آن در وظایف مختلف پردازش زبان طبیعی بسیار ساده است. در مدل پایه که توسط $\text{cite}\{\text{chen:2019}\}$ معرفی شد، یک لایه‌ی تماماً متصل بر روی هریک از خروجی‌های $\text{Ir}\{\text{BERT}\}$ قرار گرفت تا خروجی آن که به بعد آن، همان ابعاد تعبیه‌ی است، تبدیل به تعداد کلاس‌های برجسب‌ها و هدف‌های مجموعه داده شود. بدین ترتیب، خروجی توکن $\text{\$[CLS]\$}$ برای مشخص کردن هدف کاربر، و توکن‌های بخش اول ورودی به عنوان نتیجه‌ی برجسب‌زنی در نظر گرفته شد. بدین ترتیب، رمزنگار شبکه خود مدل $\text{Ir}\{\text{BERT}\}$ و رمزگشا یک لایه‌ی تماماً متصل در نظر گرفته شد.

اشتراک گذاری بردارهای توجه نیز، اطلاعات موجود وظیفه‌ی تشخیص هدف را در اختیار وظیفه‌ی تشخیص جای خالی گذاشتند. Slot-Gated از یک شبکه‌ی $\text{Ir}\{\text{LSTM}\}$ دوطرفه برای تعبیه‌ی جمله‌ی ورودی استفاده شد. این $\text{Ir}\{\text{LSTM}\}$ میان هر دو وظیفه مشترک و به عنوان رمزنگار مدل به کار گرفته شد. سپس در رمزگشا، مکانیزم توجه $\text{cite}\{\text{attention_bahdanau}\}$ برای هر وظیفه به صورت جداگانه روی آن لایه‌ی $\text{Ir}\{\text{LSTM}\}$ اعمال گردید. برای تولید احتمالات تشخیص هدف، آخرین خروجی $\text{Ir}\{\text{LSTM}\}$ همراه با بردار توجه جمع شده، سپس به یک لایه‌ی تماماً متصل تغذیه شد و با استفاده از سافت‌مکس، احتمالات خروجی برای تشخیص هدف به دست آمد. --- اما برای وظیفه‌ی تشخیص جای خالی، یک دروازه‌ی جدید محاسبه شد که درواقع یک امتیاز بود که در بردار توجه جای خالی ضرب می‌شد. نتیجه‌ی ضرب، با بردار حالت مخفی برای هر مرحله از تولید برجسب جمع می‌شد و سپس بردار نهایی بدست آمده با استفاده از یک لایه‌ی تماماً متصل و سافت‌مکس، احتمالات خروجی را تولید می‌کرد. به عبارتی، در معماری مدل $\text{Ir}\{\text{Slot-Gated}\}$ ، هیچ شبکه‌ی بازگشتی در رمزگشا استفاده نشد. برای به دست آوردن خروجی آن دروازه، جمع وزن‌دار بردار توجه تشخیص هدف، که در بعد زمانی گسترده شده بود، با بردار توجه جای خالی بدست می‌آمد.

لازم به ذکر است که معمولاً مکانیزم توجه از یک ماتریس $\text{\$}\mathbb{R}^{i \times j}\$$ تشکیل شده که در آن $\text{\$i\$}$ به عنوان طول جمله مبدا و $\text{\$j\$}$ طول جمله‌ی مقصد است؛ اما چون در تشخیص هدف تنها یک مقصد داریم، یک بردار $\text{\$}\mathbb{R}^i\$$ تولید می‌شود.

مدت‌ها است که مکانیزم توجه به همراه LSTM مورد استفاده قرار می‌گیرد. ترکیب این دو در وظیفه‌ی درک زبان طبیعی نیز مورد استفاده قرار گرفته و در زمان خود بهترین عملکرد را در پی داشت. از طرفی، به علت رابطه‌ی یک به یک واژه‌های ورودی به برجسب‌های خروجی، نیازی به فضای باز احتمالاتی خروجی یک LSTM نبود. به عبارت دیگر، نیاز نبود طول جمله‌ی خروجی پویا باشد؛ بلکه خروجی با طول ثابت و هم اندازه با ورودی مورد انتظار است. از طرف دیگر، اینکه می‌دانیم دقیقاً کدام موقعیت از ورودی، به خروجی مربوط است، خود یک مزیت محسوب می‌شود و باید از آن در طراحی معماری شبکه استفاده کرد. این اتفاق در $\text{cite}\{\text{aligned_lstm_atten_nlu}\}$ افتاد. در معماری آن مدل، برای رمزنگار، از یک LSTM دوطرفه که به صورت مشترک در دو رمزگشا مورد بهره‌وری قرار می‌گرفت. در رمزگشای تشخیص هدف، از بردار توجه هدف، و همچنین آخرین حالت مخفی LSTM رمزنگار برای کلاس بندی اهداف بهره برده شده بود. اما در رمزگشای تشخیص جای خالی، یک LSTM یک طرفه تعریف شده و از رابطه‌ی یک به یک ورودی و خروجی استفاده شد. به این منظور، در هر گام زمانی، علاوه بر بردار توجه، بردار حالت مخفی مربوط به گام زمانی ورودی که متناظر با گام زمانی خروجی بود، به مدل تغذیه می‌شد. با این ترتیب از ورودی‌ها، آموزش دو وظیفه‌ی تشخیص هدف و تشخیص جای خالی، به صورت همزمان و اشتراک آن از طریق مشترک بودن تابع خطای آن دو در تکمیل کار معماری LSTM تراز شده $\text{cite}\{\text{aligned_lstm_atten_nlu}\}$ ، با شبکه‌ی عصبی کانولوشنی و ساختار توالی ویژگی پنجره، نتایج بهتری به دست آمد $\text{cite}\{\text{Wang:18}\}$. در این شبکه که آن را Aligned CNN-BLSTM نامیدند $\text{cite}\{\text{Wang:18}\}$ ، در بخش رمزنگار، عملیات کانولوشن با اندازه هسته‌های مختلف صورت گرفته، و نتیجه‌ی حاصل پس از ترانهاد شدن، به یکدیگر الحاق می‌شد. این ساختار را توالی ویژگی پنجره نامیدند. سپس خروجی این ساختار به یک LSTM دوطرفه برای تعبیه داده می‌شد. شکل -- ساختار این شبکه را نمایش می‌دهد. ما این ساختار را به عنوان پایه‌ی کار خود در نظر گرفته ایم؛ از این رو در بخش ۴، شباهت زیادی میان ساختار رمزنگار ما و این مدل خواهید دید. بخش رمزنگار این شبکه نیز کاملاً با شبکه‌ی $\text{cite}\{\text{aligned_lstm_atten_nlu}\}$ یکسان بود. متعاقباً، اشتراک آموزش دو وظیفه نیز با اشتراک گذاری تابع خطا اتفاق می‌افتاد.

تعبیه‌ی سطح الفبا را برای درک زبان طبیعی پیشنهاد کرد. به این منظور، ابتدا کلمات را به الفای سازنده شکسته، و حروف الفبای کلمه را به یک LSTM دوطرفه دادند تا تعبیه‌ی سطح الفبای دوطرفه را ایجاد کند. در پایان با الحاق تعبیه‌ی ایجاد شده از LSTM با یک تعبیه‌ی سطح واژه مانند Word2Vec ، تعبیه‌ی نهایی واژه را ساختند. در معماری این مدل، از LSTM دو طرفه برای رمزنگار و از CRF برای رمزگشا استفاده شد.

همچنین در $\text{cite}\{\text{Siddhant:2019}\}$ از ELMo به عنوان تعبیه‌ی اولیه، LSTM دو طرفه به عنوان رمزنگار و CRF به عنوان رمزگشا استفاده کرد. در این مقاله، به ارائه یک مدل زبانی تنظیم شده برای وظیفه‌ی برجسب زنی پرداخته شد و چند استراتژی برای سبک کردن مدل زبانی ELMo ارائه شد.

از طرفی، $\text{cite}\{\text{huang:2020}\}$ $\text{Federated-Learning}$ رمزنگار چند نما را معرفی کرد. در این رمزنگار، یک رمزنگار برای تعبیه‌ی موقعیت، یک رمزنگار برای تعبیه‌ی اطلاعات محلی، یک تعبیه برای اطلاعات کلی (در سطح جمله) و یک رمزنگار اطلاعات توالی (سری زمانی) استفاده شد. از طرف دیگر، در برای آموزش مدل، یادگیری فدرال معرفی شد. در این شیوه‌ی یادگیری، یک رمزگذار بین دو مجموعه داده‌ی ATIS و SNIPS مشترک بود، به این ترتیب، برای هر مجموعه داده، به صورت جدا، رمزگشای تشخیص هدف و تشخیص جای خالی تعریف شد. در نهایت، مدل باید پارامترهای رمزنگار را میان ۴ وظیفه‌ی مختلف (دو وظیفه برای هر مجموعه داده) به اشتراک می‌گذاشت. نکته‌ی پایانی درمورد مدل $\text{Federated-Learning}$ ، استفاده‌ی آن از BERT به عنوان تعبیه‌ی کلمات بود.

معماری $\text{cite}\{\text{Qin:2021}\}$ $\text{Co-Interactive Transformer}$ برای ادغام پیش زمینه‌ی تشخیص هدف و تشخیص جای خالی در یکدیگر معرفی شد و از مکانیزم توجه چند سر transformer استفاده کرد. در معماری Co-Interactive از رمزنگار - رمزگشا، به رمزنگار - ماژول تعاملی - رمزگشا تغییر کرد. در رمزنگار از یک LSTM دوطرفه استفاده شد. سپس با استفاده از مکانیزم توجه تعریف شده در برای تشخیص هدف و تشخیص جای خالی، به صورت جداگانه ماتریس توجه به دست آمد. در ماژول تعاملی پیشنهادی، از مکانیزم توجه ترنسفورمر استفاده شد؛ به این نحو که به ازای هر وظیفه‌ی، یک تبدیل خطی از ماتریس توجه آن وظیفه برای کلید، مقدار و پرسش ایجاد شد. سپس مکانیزم توجه برای هر وظیفه، با استفاده از کلید و مقدار همان وظیفه و پرسش وظیفه‌ی دیگر، اعمال شد. بدین ترتیب در این ماژول، اطلاعات ماتریس توجه مربوط به هر دو وظیفه با یکدیگر ترکیب شد. در گام بعدی، مانند معماری ترنسفورمر، ماتریس ورودی با خروجی مکانیزم توجه توصیفی برای هر وظیفه، جمع و نرمالایز شد. در لایه‌ی تغذیه به جلوی ماژول تعاملی، خروجی هر کدام از داده‌های نرمالایز شده مربوط به وظایف با یکدیگر الحاق شده و به یک لایه‌ی تغذیه به جلو داده می‌شود. حاصل خروجی این لایه‌ی تغذیه به جلو با خروجی لایه‌ی توجه تعاملی، جمع و نرمالایز شد. در پایان، برای رمزگشای تشخیص جای خالی از CRF ، و برای رمزگشای تشخیص هدف از Max - Pooling استفاده شد.

در تلاشی دیگر برای ادغام بهینه‌ی دو وظیفه‌ی تشخیص هدف و تشخیص جای خالی، \cite{Yang:2021}\{AISE\} معرفی شد. \{AISE\} \{Attending to Intent and Slots Explicitly\} متشکل از یک رمزنگار مشترک، یک رمزنگار تشخیص هدف و مکانیزم پیشنهادی توجه چند سر پوشیده‌ی آگاه از موقعیت بود. مدل \{AISE\} بر مبنای \{LSTM\} و مکانیزم توجه چند سر بنا شد. در رمزگشای تشخیص هدف، از مکانیزم ادغام توجه چند سر استفاده کردند. ادغام توجه چند سر، نوعی از ادغام میانگین است که در آن وزن‌ها در میانگین‌گیری قابل تنظیم است. در این مکانیزم، کلید و مقدار از تعبیه‌ی رمزنگار و پرسش یک بردار، با مقادیر اولیه تصادفی و قابل آموزش است. خروجی این مکانیزم نیز، یک بردار با بعد تعبیه‌ی ورودی می‌باشد. در \{AISE\}، برای رمزگشای تشخیص جای خالی، مکانیزم چند سر پوشیده‌ی آگاه از موقعیت معرفی شد. در این مکانیزم، از یک مکانیزم چند سر پوشیده و برای تعبیه‌سازی ورودی نیز از تعبیه‌ی نسبی \cite{relative_positioning_transformer}\{AISE\} به صورت ترتیبی تولید شدند، در مکانیزم چند سر پوشیده، از تعبیه‌ی رمزنگار در موقعیت مربوط به نشانه‌ی درحال تولید به عنوان بردار پرسش استفاده می‌شد. همچنین از الحاق بردار زمینه‌ی تشخیص هدف و ماتریس تعبیه‌ی ابتدایی جمله به عنوان کلید و مقدار استفاده شد. در پایان، برجسب تولیدی در هر موقعیت، با استفاده از الحاق و سافت‌مکس بردار خروجی توجه چند سر پوشیده با بردار تعبیه‌ی مربوط به آن نشانه به دست آمد. مدل \{PKJL\} \{Prior Knowledge Joint Learning\} برای افزایش به اشتراک گذاری اطلاعات قبلی با دو وظیفه‌ی تشخیص جای خالی و تشخیص هدف معرفی شد. منظور از اطلاعات قبلی، احتمال وقوع هر برجسب هدف با هر برجسب جای خالی در یک نمونه‌ی آموزشی است. مدل \{PKJL\} با محاسبه‌ی این احتمالات و تغذیه‌ی آن به مدل خود در یک لایه‌ی ابداعی، دانش قبلی را با مدل به اشتراک گذاشت. معماری این مدل از یک \{LSTM\} دو طرفه در رمزنگار و بردار احتمالات وقوع هر برجسب با هر هدف تشکیل می‌شد. یک ماژول ابداعی به نام لایه‌ی یکپارچه سازی اطلاعات بعد از رمزنگار ایجاد شد که سه ماتریس توجه برای تشخیص هدف، تشخیص جای خالی و دانش قبلی در آن ایجاد می‌شدند. این سه ماتریس با یکدیگر تجمیع و تبدیل خطی شدند تا خروجی ماژول ابداعی تکمیل شود. خروجی این ماژول، برای رمزنگار تشخیص جای خالی، که یک \{CRF\} و رمزنگار تشخیص هدف که یک لایه‌ی خطی همراه با سافت‌مکس بود، استفاده

نام فایل	نام مدل	ساختار	نقش مدل زبانی
----------	---------	--------	---------------

wang2020.pdf	SASGBC	Bert as enc	برت- اسلات گیت- سلف اتنشن-سی آراف
--------------	--------	-------------	-----------------------------------

1907.00390.pdf	Interrelated	iteration-crf	بای الاستیم،
----------------	--------------	---------------	--------------

1902.10909.pdf	Bert base	encoder	خود برت، توکن ها جای توکن و اینتنت جای سی ال
----------------	-----------	---------	--

NAACL18_SlotGated.pdf	Slot gated	bilstm-2 context vectors bandanahu, slot gate mechanism with ff network and softmax	
-----------------------	------------	---	--

1352.PDF	LSTM ATTEN	LSTM with ATTEN	
11.pdf	cnn-aligned-blstm	cnn-wfs-lstm-lstmout	
daha-fatima-zohra-deep-r	character+BILSTM	CRF	
	Elmo-blstm+crf	Elmo-blstm+crf	embedding
2020.coling-main.310.pdf	Federated Learning		embedding
2010.03880.pdf	Co-interactive Transformer		embedding

yang2020.pdf	AISE	Lstm enc - Multihead atten for Intent Proposed Positionawared Multihead SF	embedding
--------------	------	---	-----------

9789811269264_0001.p df	Prior Knowledge	LSTM	Nadare
----------------------------	-----------------	------	--------

١٢

٤

١١

٣

1

2

5