



دانشگاه صنعتی امیرکبیر  
(پلی تکنیک تهران)  
دانشکده ریاضی و علوم کامپیوتر

درس هوش مصنوعی و کارگاه

گزارش ۲: NEUROLOGIC A\* esque Decoding  
**Constrained Text Generation with Lookahead Heuristics**

نگارش

کیارش مختاری دیزجی

۹۸۳۰۰۳۲

استاد اول

دکتر مهدی قطعی

استاد دوم

بهنام یوسفی مهر

فروردین ۱۴۰۲

## چکیده

پارادایم غالب برای تولید متن با شبکه عصبی، رمزگشایی از چپ به راست با استفاده از مدل‌های زبان خودرگرسیون<sup>۱</sup> است. با این حال، تولید با قید یا تولید قابل کنترل تحت محدودیت‌های واژگانی پیچیده، به آینده‌نگری و برنامه‌ریزی مسیرهای امکان پذیر آینده نیاز دارد.

در این مقاله با الهام گیری از الگوریتم  $A^*$ ، یک الگوریتم رمزگشایی به نام NEUROLOGIC  $A^*$ esque که شامل پیش‌بینی هیوریستیک از future cost می‌باشد، ارائه شده است و هیوریستیک‌های پیش‌بینی کارآمدی را توسعه داده است که برای مدل‌های زبانی در مقیاس بزرگ<sup>۲</sup> کارآمد هستند، و روشش را جایگزینی برای تکنیک‌های رایجی مانند beam search و top-k sampling می‌داند. برای تولید با قید<sup>۳</sup>، با استفاده از رمزگشایی NEUROLOGIC و استفاده از ویژگی انعطاف‌پذیری‌اش، آن را با قیدهای منطقی که با پیش‌بینی‌های  $A^*$  برای ارضای قیدهای آینده است، ترکیب کرده است.

این رویکرد به عملکرد پیشرفته‌ای در زمینه تولید متن جدولی<sup>۴</sup>، ترجمه ماشینی با قید<sup>۵</sup>، و تولید با قید کلمه کلیدی<sup>۶</sup> دست یافته است.  $A^*$  NEUROLOGIC قدرت رمزگشایی را برای بهبود و فعال کردن قابلیت‌های جدید LLM‌ها نشان می‌دهد.

## واژه‌های کلیدی:

هیوریستیک، NEUROLOGIC decoding، الگوریتم  $A^*$ ، مدل‌های زبانی در مقیاس بزرگ

<sup>1</sup> autoregressive

<sup>2</sup> Large-scale Language Model(LLM)

<sup>3</sup> Constrained generation

<sup>4</sup> Table-to-text generation

<sup>5</sup> Constrained machine translation

<sup>6</sup> Keyword-constrained generation

صفحه	فهرست مطالب
أ	چکیده.....
۱	۱. فصل اول مقدمه.....
۴	۲. فصل دوم NEUROLOGIC A* Decoding.....
۵	۲-۱ - رمزگشایی با استفاده از رمزگشایی A* lookahead.....
۷	۲-۲ - تولید بدون قید با استفاده از NEUROLOGIC*.....
۷	۲-۳ - تولید با قید با استفاده از NEUROLOGIC*.....
۸	۳. فصل سوم آزمایشات: تولید با قید.....
۹	۳-۱ - تولید قضاوت صحیح با قید.....
۱۰	۳-۲ - ترجمه ماشینی با قید.....
۱۰	۳-۳ - تولید جدول - به - متن.....
۱۱	۳-۴ - تولید سوال با قید.....
۱۲	۴. فصل چهارم آزمایشات: تولید بدون قید.....
۱۳	۴-۱ - تولید داستان با مفاهیم متعارف.....
۱۴	۵. فصل پنجم جمع‌بندی و نتیجه‌گیری.....
۱۵	منابع.....

۱- نمونه‌ای از کاربرد الگوریتم NEUROLOGIC A*	۲
۲- اجرای روش‌های رمزگشایی مختلف با GPT-2 نظارت‌شده یا خارج از قفسه در مجموعه آزمایشی	
COMMONGEN	۹
۳- نتایج مربوط به ترجمه ماشینی با قید	۱۰
۴- عملکرد روش‌های مختلف رمزگشایی برای تولید جدول به متن	۱۰
۵- عملکرد الگوریتم‌های رمزگشایی بدون نظارت مختلف در تولید سؤال	۱۱
۶- عملکرد الگوریتم‌های رمزگشایی مختلف در دادگان RocStories	۱۳

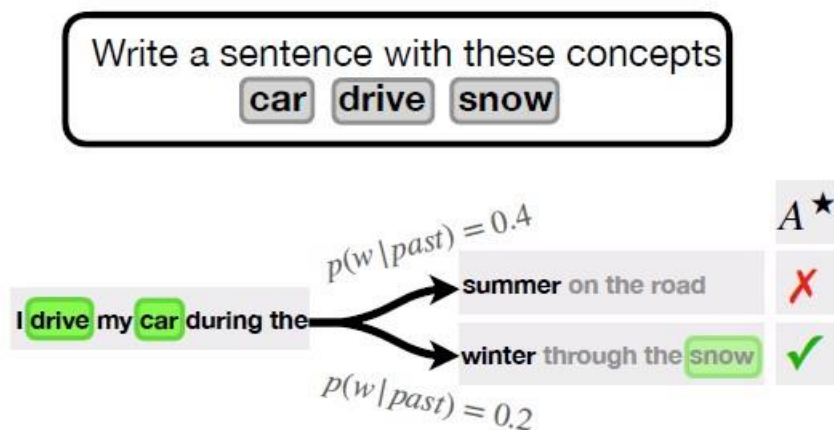
## ۱. فصل اول

### مقدمه

پارادایم غالب برای تولید متن با شبکه عصبی، رمزگشایی از چپ به راست با استفاده از مدل‌های زبانی خودرگرسیون مانند GPT-2/3 می‌باشد و با توجه به این پارادایم روش‌های رمزگشایی معمول مانند beam search یا top-k/p sampling مشخص می‌کنند چه توکنی برای مرحله بعد با توجه به وقایعی که در گذشته اتفاق افتاده است و همینطور بدون نگاه به اینکه در آینده چه اتفاقی افتاده است، ساخته بشود. در حالی که این فقدان آینده نگری اغلب در تولید متن با قید کافی است، برنامه‌ریزی از قبل برای ترکیب تمام محتواهای مورد نظر در خروجی تولید شده بسیار مهم است.

می‌توان با استفاده از الگوریتم‌های جستجو مانند  $A^*$ ، چالش برنامه ریزی برای آینده را با استفاده از پیش‌بینی هیوریستیک از future cost در هنگام تصمیم‌گیری برطرف کرد. بنابراین با استفاده از این الگوریتم و الگوریتم رمزگشایی برای تولید متن با استفاده از شبکه عصبی مانند beam search، مدلی به نام NEUROLOGIC  $A^*$  ایجاد شده است که کارآمدی لازم را برای LLMها فراهم می‌کند.

این مقاله یک lookahead هیوریستیک توسعه داده است که هزینه تقریبی در هر مرحله رمزگشایی را براساس ادامه دنباله پیش‌بینی می‌کند.



۱- نمونه‌ای از کاربرد الگوریتم NEUROLOGIC  $A^*$

همانطور که از تصویر ۱-۱ مشخص است اگرچه summer با توجه به کلمات که قبلاً تولید شده‌اند، احتمالاً کلمه بعدی است، اما \*NEUROLOGIC به آینده نگاه می‌کند و به همین دلیل انتخاب winter منجر می‌شود که در آینده احتمال را بالاتر ببرد. بنابراین، winter با وجود احتمال کمتری نسبت به summer ترجیح داده می‌شود.

با توجه به آزمایشات گسترده حول five generation task نشان می‌دهد که این الگوریتم بسیار خوب عمل می‌کند همچنین این الگوریتم می‌تواند نیاز به مجموعه داده‌های پرهزینه را که به صورت دستی برای نظارت<sup>۷</sup> صریح حاشیه نویسی می‌شوند، کاهش دهد.

به طور خلاصه، \*NUROLOGIC A یک الگوریتم رمزگشایی جدید برای تولید متن موثر و کارآمد است. طبق گفته مقاله، این اولین الگوریتم \*A-مانند برای تولید متن هدایت‌شده از طریق هیوریستیک‌های پیش‌رو است. این الگوریتم همه‌کاره است، زیرا می‌توان آن را از طریق قیدهای استنتاج-زمان<sup>۸</sup> برای انواع وظایف اعمال کرد و نیاز به داده‌های برچسب‌گذاری شده پرهزینه را کاهش داد. آزمایش‌های گسترده اثربخشی آن را بر روی چندین معیار مهم نشان می‌دهد.

---

<sup>7</sup> supervised

<sup>8</sup> Inference-time constraint

۲. فصل دوم

**NEUROLOGIC A\* Decoding**



## ۲-۱- رمزگشایی با استفاده از رمزگشایی A\* lookahead

تولید توالی دنباله‌ها به صورت تولید یک دنباله خروجی  $y$  با توجه به یک دنباله ورودی  $x$  است. با در نظر گرفتن مدل‌های استاندارد از چپ به راست و خودرگرسیو،  $p_\theta(y|x) = \prod_{t=1}^{|y|} p_\theta(y_t | y_{<t}, x)$  و با حذف  $x$  برای کاهش پراکندگی و شلوغی رمزگشایی شامل حل معادله،

$$y_* = \arg \max_{y \in \mathcal{Y}} F(y)$$

۱

که  $\mathcal{Y}$  مجموعه همه دنباله‌ها می‌باشد. در این محیط  $F(y)$  به شکل  $s(y) + H(y)$  است، که  $s(y)$  نیز به شکل  $\log p_\theta(y)$  است و  $H(y)$  نیز یا صفر است یا مقدار امتیازی است که برای ارضای قیدهای روی  $y$  بدست آمده است. لازم به ذکر است که دیدگاه رمزگشایی‌اش براساس جستجو گسسته می‌باشد که حالت‌ها پیشوندهای جزئی هستند،  $y_{<t}$ ، اکشن‌ها همان توکن‌های داخل مجموعه واژگان  $\mathcal{V}$  ( $y_t \in \mathcal{V}$ ) و ترنزیشن‌ها یک توکن را به پیشوند اضافه می‌کنند،  $y_{<t} \circ y_t$ . هر مرحله رمزگشایی شامل:

(۱) گسترش مجموعه کاندیداهای استتیت بعدی.

(۲) امتیاز دهی هر کاندید.

(۳) انتخاب  $k$  تا از بهترین کاندیدها.

$$Y'_t = \{y_{<t} \circ y_t \mid y_{<t} \in Y_{t-1}, y_t \in \mathcal{V}\},$$

$$Y_t = \underset{(y_{<t}, y_t) \in Y'_t}{\arg \text{topk}} \{f(y_{<t}, y_t)\}$$

۲

Lookahead هیوریستیک. این روش تخمینی از آینده را در انتخاب کاندید در خود گنجانده است. در حالت ایده‌آل، ما می‌خواهیم کاندیدهایی را انتخاب کنیم که در مسیرهای بهینه هستند و معادله ۲ را با معادله زیر جابجا می‌کنیم:

$$Y_t = \underset{(y_{<t}, y_t) \in Y'_t}{\arg \text{topk}} \left\{ \max_{y_{>t}} F(y_{<t}, y_t, y_{>t}) \right\}.$$

۳

۵

اما محاسبه معادله ۳، دو مشکل دارد:

(۱) تابع هدف  $F(y)$  یا ناشناخته می‌تواند باشد و یا محاسباتش پیچیده باشد.

(۲) فضای مسیرهای آینده  $y_{<t}$  بسیار بزرگ است.

بنابراین تابع هدف را با استفاده از یک هیوریستیک سبک وزن<sup>۹</sup>  $h(\cdot)$  تقریب می‌زنیم:

$$Y_t = \arg \text{topk}_{(y < t, y_t) \in Y'_t} \left\{ s(y_{\leq t}) + \max_{y > t} h(y_{<t}, y_t, y_{>t}) \right\},$$

۴

که  $s(y_{\leq t}) = \log p_\theta(y_{\leq t})$  است. برای اینکه جستجو قابل انجام باشد، مجموعه‌ای از ادامه‌های

lookahead را جستجو می‌کنیم بنابراین معادله‌ی ۳ تقریب زده به صورت زیر است:

$$Y_t = \arg \text{topk}_{y_{\leq t} \in Y'_t} \left\{ s(y_{\leq t}) + \max_{\mathcal{L}_\ell(y_{\leq t})} h(y_{\leq t+\ell}) \right\},$$

۵

که در اینجا هر عنصر  $y_{t+1:t+\ell}$  از مجموعه  $\mathcal{L}_\ell(y_{\leq t})$  همان طول ( $\ell$ ) ادامه دنباله  $y_{\leq t}$  است. Beam search متناظر با این است که  $h$  و  $\ell$  را صفر قرار دهیم.

روش مقاله و A\* search و beam search همگی یک کلاس کلی از الگوریتم‌ها هستند که براساس

موارد زیر متفاوت می‌باشند:

(۱) کدام کاندید گسترش می‌یابد.

(۲) کدام کاندید هرس می‌شود.

(۳) و چگونه کاندیدها امتیاز دهی می‌شوند.

حال مزایای عملی گسترش و هرس به سبک beam search را به همراه هیوریستیک‌های A\* برای تخمین آینده ترکیب می‌کند و الگوریتم A\*esque decoding را ارائه می‌کند. همچنین مقاله روش‌های

<sup>۹</sup> Lightweight heuristic

بسیاری را برای تولید lookahead های  $\mathcal{L}_\ell(y_{\leq t})$  مقایسه می کند و در آخر روش خودش را برای حالت های بدون شرط و با شرط ارائه می کند.

## ۲-۲- توليد بدون قيد با استفاده از NEUROLOGIC\*

ابتدا یک تنظیم رمزگشایی استاندارد را در نظر می گیریم:

$$\arg \max_{y \in \mathcal{Y}} \log p_\theta(y | x)$$

کاندیدها را بر اساس ترکیبی از تاریخچه<sup>۱۰</sup> و آینده تخمین زده<sup>۱۱</sup>، با استفاده از احتمال lookahead به عنوان یک هیوریستیک امتیاز می دهیم. یعنی در مرحله  $t$ ام رمزگشایی از معادله ۵ استفاده می کنیم:

$$h(y_{\leq t+\ell}) = \lambda \log p_\theta(y_{t+1:t+\ell} | y_{\leq t}, x)$$

۶

که  $\lambda$  در اینجا کنترل کننده میزان اعتماد به تاریخچه در مقابل آینده تخمین زده شده است.

## ۲-۳- توليد با قيد با استفاده از NEUROLOGIC\*

روش مقاله که بر روی تابع امتیازدهی NEUROLOGIC با برآورد رضایت شرطهای آینده بهبود می یابد و چیزی که به آن اضافی می شود یک هیوریستیک lookahead است که امتیاز کاندید را متناسب با احتمال برآورده کردن شرطهای اضافی در lookahead  $y_{t+1:t+\ell}$  تنظیم می کند:

$$h_{\text{future}}(y_{\leq t+\ell}) = \lambda_2 \max_{D(a, y_{\leq t})} \log p_\theta(D(a, y_{t+1:t+\ell}) | x, y_{\leq t}),$$

۷

<sup>10</sup> History

<sup>11</sup> Estimated future

### ۳. فصل سوم

#### آزمایشات: تولید با قید

۳-۱- تولید قضاوت صحیح<sup>۱</sup> با قید

Decode Method	Automatic Evaluation						Human Evaluation			
	ROUGE-L	BLEU-4	METEOR	CIDEr	SPICE	Coverage	Quality	Plausibility	Concepts	Overall
<i>Supervised</i>										
CBS (Anderson et al., 2017)	38.8	20.6	28.5	12.9	27.1	97.6	2.27	2.35	2.51	2.23
GBS (Hokamp and Liu, 2017)	38.2	18.4	26.7	11.7	26.1	97.4	2.06	2.17	2.29	2.01
DBA (Post and Vilar, 2018a)	38.3	18.7	27.7	12.4	26.3	97.5	2.23	2.30	2.43	2.15
NEUROLOGIC (Lu et al., 2021)	42.8	26.7	30.2	14.7	30.3	97.7	2.54	2.56	2.67	2.50
NEUROLOGIC <sup>*</sup> (greedy)	43.6	28.2	30.8	15.2	30.8	97.8	2.66	2.67	2.73	2.59
NEUROLOGIC <sup>*</sup> (sample)	43.4	27.9	30.8	15.3	31.0	97.7	2.64	2.64	2.74	2.58
NEUROLOGIC <sup>*</sup> (beam)	43.2	28.2	30.7	15.2	31.0	97.6	2.68	2.67	2.76	2.60
<i>Unsupervised</i>										
TSMH (Zhang et al., 2020)	24.7	2.2	14.5	3.6	15.4	71.5	1.85	1.92	1.95	1.63
NEUROLOGIC (Lu et al., 2021)	41.9	24.7	29.5	14.4	27.5	96.7	2.64	2.52	2.68	2.50
NEUROLOGIC <sup>*</sup> (greedy)	44.3	28.6	30.7	15.6	29.6	97.1	2.78	2.70	2.77	2.70

۲ - اجرای روش‌های رمزگشایی مختلف با GPT-2 نظارت‌شده یا خارج از قفسه در مجموعه آزمایشی COMMONGEN

با توجه به جدول بالا می‌توان نتایج زیر را از آزمایش روش‌های مختلف رمزگشایی همراه با قید گرفت:

(۱) الگوریتم ارائه شده از دیگر الگوریتم‌های رمزگشایی با قید بهتر عمل کرده است و نکته قابل توجه این است که NEUROLOGIC<sup>\*</sup> بدون نظارت<sup>۲</sup> از همه روش‌های نظارت<sup>۳</sup> شده بر اساس ارزیابی انسانی بهتر عمل می‌کند.

(۲) NEUROLOGIC<sup>\*</sup> کیفیت تولید را بهبود می‌بخشد و در عین حال با رضایت بالایی قیدها را حفظ می‌کند. این تفاوت به ویژه در مورد صفر شات<sup>۴</sup> قابل توجه است، جایی که به دلیل عدم نظارت و فضای خروجی زیاد، فضای بیشتری برای ترکیب سیگنال‌های مبتنی بر قید وجود دارد.

(۳) در ارزیابی با انسان می‌توان دید که در میان تابع‌های هیوریستیک lookahead برای NEUROLOGIC<sup>\*</sup> beam، بهترین عملکرد را داشته و greedy نیز ضعیف‌ترین عملکرد را داشته است.

<sup>1</sup> Commensense

<sup>2</sup> Unsupervised

<sup>3</sup> Supervised

<sup>4</sup> Zero-shot

## ۳-۲- ترجمه ماشینی با قید

Method	Dinu et al.		Marian MT	
	BLUE	Term%	BLUE	Term%
Unconstrained	25.8	76.3	32.9	85.0
train-by-app.	26.0	92.9	—	—
train-by-rep.	26.0	94.5	—	—
Post and Vilar (2018a)	25.3	82.0	33.0	94.3
NEUROLOGIC	26.5	95.1	33.4	97.1
NEUROLOGIC <sup>*</sup> (greedy)	26.7	95.8	33.7	97.2
NEUROLOGIC <sup>*</sup> (sample)	26.6	95.4	33.7	97.2
NEUROLOGIC <sup>*</sup> (beam)	26.6	95.8	33.6	97.2

# T	# Sents.	Decode Method	BLEU	Term%
1	378	Beam search	25.4	79.6
		NEUROLOGIC	26.2	95.2
		NEUROLOGIC <sup>*</sup>	26.3	95.8
2+	36	Beam search	28.1	85.0
		NEUROLOGIC	28.9	93.7
		NEUROLOGIC <sup>*</sup>	29.3	96.5

۳- نتایج مربوط به ترجمه ماشینی با قید.

با آزمایش الگوریتم NEUROLOGIC<sup>\*</sup> در این حالت نیز از همه روش‌های قبلی هم در BLEU و هم در term coverage عملکرد بهتری دارد و همچنین می‌توان دید که NEUROLOGIC<sup>\*</sup> در حالت‌های که قید با پیچیدگی زیاد داریم نسبت به روش‌های دیگر بهتر عمل می‌کند.

## ۳-۳- تولید جدول - به - متن

Decode Method	NIST	BLEU	METEOR	CIDEr	ROUGE	Coverage
Beam Search	3.82	42.8	32.6	10.8	57.8	73.6
CBS	6.50	42.3	36.4	13.0	54.3	91.6
GBS	6.26	40.7	36.7	12.9	54.2	94.1
NEUROLOGIC	6.95	47.6	38.9	16.3	58.7	97.6
NEUROLOGIC <sup>*</sup> (greedy)	7.11	49.2	40.0	17.5	60.0	100.0
NEUROLOGIC <sup>*</sup> (beam)	7.01	48.9	40.0	17.2	59.8	99.9
NEUROLOGIC <sup>*</sup> (sample)	7.11	49.3	40.1	17.5	60.0	100.0

۴- عملکرد روش‌های مختلف رمزگشایی برای تولید جدول به متن

NEUROLOGIC<sup>\*</sup> به طور قابل ملاحظه‌ای از همه روش‌های قبلی با توجه به تمام معیارها، بهتر عمل می‌کند و به طور مداوم کیفیت تولید را بهبود می‌بخشد در حالی که تقریباً رضایت کامل از قیدها را بدست می‌آورد. در کارهایی مانند CBS و GBS می‌توان دید بهبود ارضای قیود را اما با این حال بروی کیفیت متن‌های تولید شده تاثیر منفی گذاشته است.

## ۳-۴- تولید سوال با قید

Decode Method	Automatic Evaluation						Human Evaluation			
	ROUGE	BLEU	METEOR	CIDEr	SPICE	Coverage	Grammar	Fluency	Meaningfulness	Overall
CGMH (Miao et al., 2019)	28.8	2.0	18.0	5.5	21.5	18.3	2.28	2.34	2.11	2.02
TSMH (Zhang et al., 2020)	42.0	4.3	25.9	10.4	37.7	<u>92.7</u>	2.35	2.28	2.37	2.22
NEUROLOGIC (Lu et al., 2021)	38.8	11.2	24.5	18.0	41.7	90.6	2.78	2.71	2.49	2.51
NEUROLOGIC★ (greedy)	<b>43.7</b>	<b>14.7</b>	<u>28.0</u>	<b>20.9</b>	<u>47.7</u>	<b>100.0</b>	<b>2.83</b>	<u>2.77</u>	<u>2.74</u>	<b>2.76</b>
NEUROLOGIC★ (beam)	42.9	14.4	27.8	20.3	46.9	<b>100.0</b>	<u>2.81</u>	<b>2.86</b>	<b>2.76</b>	<u>2.75</u>
NEUROLOGIC★ (sample)	<u>43.5</u>	<u>14.6</u>	<b>28.2</b>	<u>20.8</u>	<b>47.8</b>	<b>100.0</b>	<b>2.83</b>	2.75	<b>2.76</b>	2.73

۵- عملکرد الگوریتم‌های رمزگشایی بدون نظارت مختلف در تولید سؤال.

NEUROLOGIC\* از همه روشهای قبلی با توجه به معیارهای خودکار و دستی بهتر عمل می‌کند. این به طور قابل ملاحظه ای کیفیت تولید را افزایش می‌دهد و در عین حال رضایت کامل از محدودیت را به دست می‌آورد. تفاوت بین NEUROLOGIC و NEUROLOGIC\* در مقایسه با سایر وظایف بسیار زیاد است. مشکل جستجو در اینجا بسیار سخت‌تر است، به دلیل عدم نظارت و قید منطقی پیچیده که شامل کلمات کلیدی و نحو است.

#### ۴. فصل چهارم

#### آزمایشات: تولید بدون قید



## ۴-۱- تولید داستان با مفاهیم متعارف

Decode Method	Fluency			Diversity		Human Eval				
	PPL	BLEU-1	BLEU-2	Uniq. 3-gram	Uniq. 4-gram	Grammar	Fluency	Coherence	Interest	Overall
beam search	2.24	33.7	16.5	34.09k	41.91k	2.81	2.50	2.46	2.27	2.32
beam search + A*esque (greedy)	<b>2.11</b>	<u>34.3</u>	<u>16.7</u>	34.94k	43.02k	<b>2.94</b>	<u>2.71</u>	2.56	2.50	<u>2.57</u>
beam search + A*esque (beam)	<u>2.14</u>	<b>34.4</b>	<b>16.8</b>	35.03k	<u>43.12k</u>	<b>2.94</b>	<b>2.72</b>	<b>2.62</b>	<b>2.61</b>	<b>2.63</b>
beam search + A*esque (sample)	2.16	<b>34.4</b>	<u>16.7</u>	<b>35.41k</b>	<b>43.64k</b>	<u>2.92</u>	<u>2.71</u>	<u>2.59</u>	<u>2.52</u>	<u>2.57</u>
top-k sample	4.01	31.4	13.9	<u>48.36k</u>	<u>56.62k</u>	2.69	2.38	2.23	2.30	2.15
top-k sample + A*esque (greedy)	<b>3.68</b>	<u>32.1</u>	<u>14.3</u>	<b>48.44k</b>	<b>56.63k</b>	<b>2.88</b>	<b>2.57</b>	<b>2.48</b>	<b>2.49</b>	<b>2.47</b>
top-k sample + A*esque (beam)	3.75	<b>32.2</b>	<b>14.4</b>	48.27k	56.36k	<u>2.84</u>	2.49	2.39	2.40	2.34
top-k sample + A*esque (sample)	<u>3.70</u>	32.0	14.2	48.04k	56.15k	<u>2.84</u>	<u>2.55</u>	<u>2.47</u>	<u>2.48</u>	<u>2.44</u>

## ۶- عملکرد الگوریتم های رمزگشایی مختلف در دادگان RocStories.

به این صورت کار می کند که با دادن یک دستور،  $\mathcal{M}$ ، وظیفه مدل تولید ادامه داستان،  $\mathcal{Y}$ ، است و برای این کار از دادگان RocStories استفاده شده است. جدول شامل نتیجه الگوریتم های beam search و topk sampling با A\*esque و بدون آن است. می توانیم ببینیم که هیوریستیک A\*esque هم beam search و هم topk sampling را برای تولید داستان های روان، منسجم و جالب تر امکان پذیر می کند. برای beam search، A\*esque نه تنها کیفیت تولید متن را افزایش داده است اما همچنین تنوع تولید را افزایش می دهد که با تعداد n-gram های منحصر به فرد نشان داده می شود. برای topk sampling علاوه بر افزایش کیفیت تولید، تنوع مقایسه پذیری نیز حفظ شده است. لازم به ذکر است که beam lookahead برای beam search و greedy lookahead برای topk sampling بهترین نتیجه را می دهد.

## ۵. فصل پنجم

### جمع‌بندی و نتیجه‌گیری

این مقاله با استفاده از الگوریتم  $A^*$  توانست الگوریتم رمزگشایی  $NEUROLOGIC A^*$  را معرفی کند که تابع هیوریستیکی مانند  $A^*$  برای پیش‌بینی از آینده برای تولید متن از چپ به راست فراهم می‌کند. این الگوریتم باعث پیشرفت الگوریتم‌های رمزگشایی در مسائل با قید یا بدون قید شده است و همچنین روشی ارائه داده است برای این که علاوه بر نگاه کردن به گذشته به آینده نیز برای تولید هرچه بهتر متن داشته باشیم و این می‌تواند آغازی برای اینچنین الگوریتم‌ها باشد.

## منابع

Lu, Ximing, Sean Welleck, Peter West, Liwei Jiang, Jungo Kasai, Daniel Khashabi, Ronan Le Bras et al. "Neurologic a\* esque decoding: Constrained text generation with lookahead heuristics." arXiv preprint arXiv:2112.08726.(۲۰۲۱)