Diverse Keyphrase Generation with Neural Unlikelihood Training

Objectives

- Analyze issues prevalent in Seq2Seq models trained using maximum likelihood estimation
- Propose measures to address repetitions in keyphrase generation models
- Improve diversity of generated keyphrases while maintaining output quality

Motivation

- MLE Training $\rightarrow 26\%$ keyphrase level duplication
- Existing solutions \rightarrow Adhoc post-processing

Proposed Approach

- Principled solution by adopting **unlikelihood** objective to train the model
- Novel copy token unlikelihood loss
- K-step ahead token prediction to incentivize model planning
- K-step ahead unlikelihood losses

Diversity Evaluation Metrics

- % Duplicate KPs, % Duplicate Tokens
- Pairwise keyphrase similarity at token level (Self-BLEU), character level (Edit-Dist), semantic level (**Emb-Sim**)

Ground	image segmentation; region merging;
Truth	dynamic programming;
	wald sequential probability ratio test
catSeq	image segmentation; region merging;
MLE	region merging ; dynamic programming ;
Baseline	image segmentation
catSeqTG-	image segmentation; region merging;
2RF1 (RL)	dynamic programming ; <i>image segmentation</i> ;
	dynamic programming
DivKGen	image segmentation; region merging;
(UL)	region merging ; dynamic programming ;
	nearest neighbor graph
DivKGen	image segmentation ; dynamic programming ;
(Full)	region merging; stopping criterion

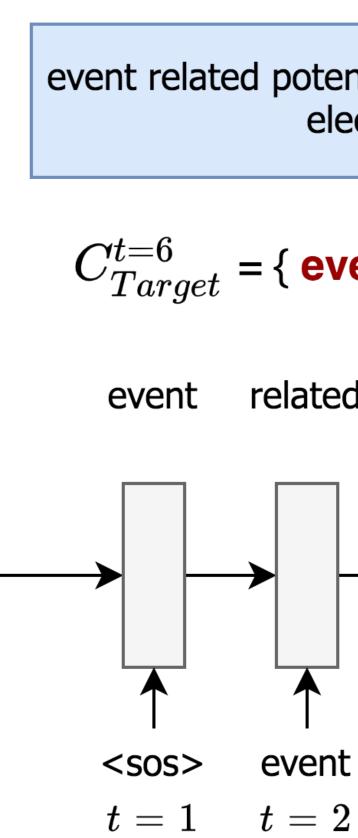
Table 1: Case Study on KP20K dataset — Article title and abstract are provided as model inputs.

Hareesh Bahuleyan and Layla El Asri Borealis AI, Montreal, Canada

(a) Target + Copy Unlikelihood Training

Source Document

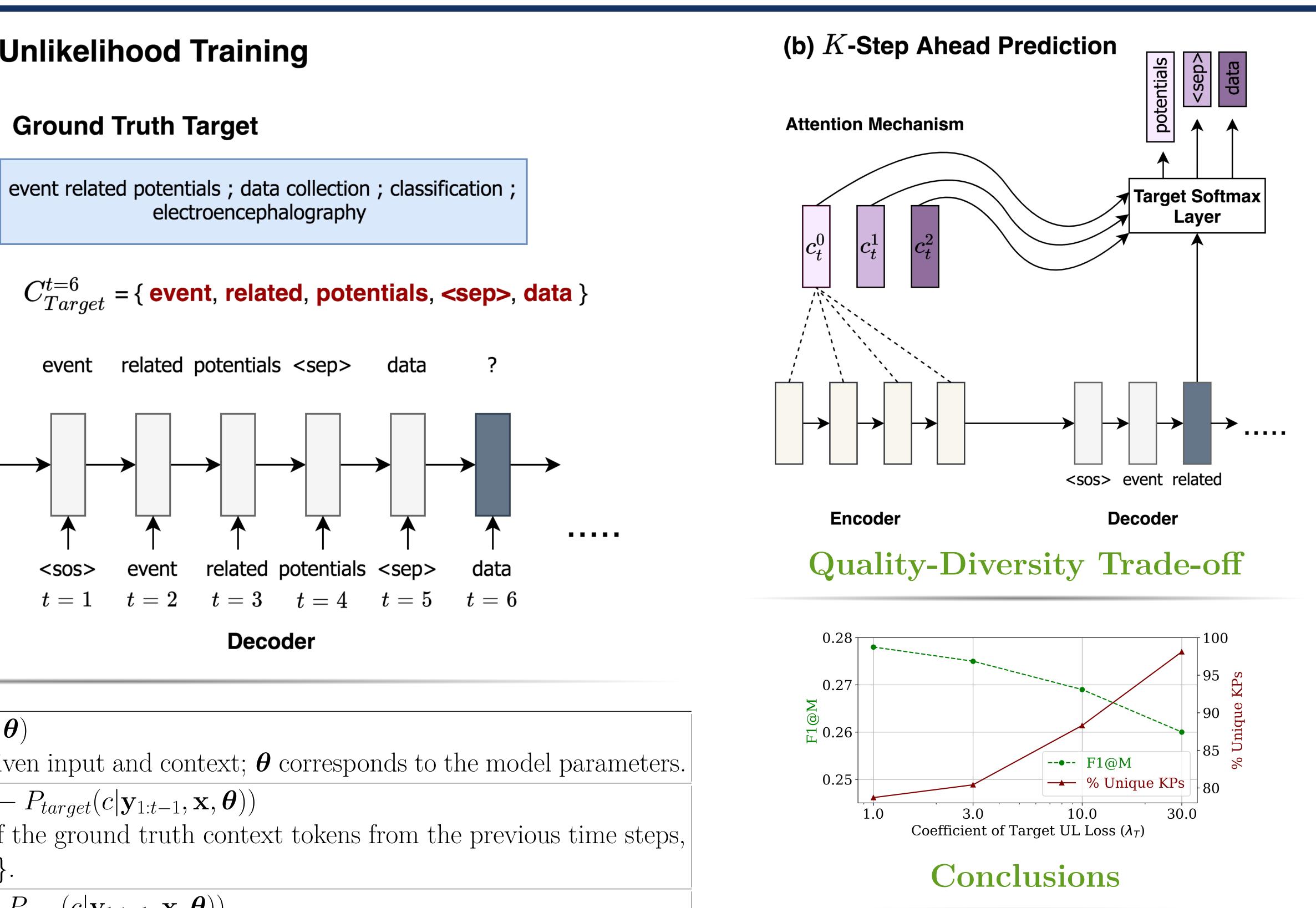
a framework for a real time intelligent and interactive brain computer interface (...) accuracy processing time comparison for detection of event related potentials . an implementation of a prototype system using the proposed bci framework . real time eeg data collection and classification of erps using hex o speller.



MLE	$ \mathcal{L}_{ ext{MLE}} = -$	$-\Sigma^L$, lo	or $P(\eta_{\mu} $	$\mathbf{V}_{1, \mathbf{J}} = \mathbf{I} \cdot \mathbf{X}$	$(\boldsymbol{\theta})$						
		v-1			· · · /	nd context: F	7 corresponds	to the model	narameters		
		-			~ -	,			parameters.		
		$\mathcal{L}_{\text{TargetUL}} = -\sum_{t=1}^{L} \sum_{c \in \mathcal{C}_{\text{Target}}} \log \left(1 - P_{target}(c \mathbf{y}_{1:t-1}, \mathbf{x}, \boldsymbol{\theta}) \right)$									
		Negative candidate list consists of the ground truth context tokens from the previous time steps,									
	i.e., $\mathcal{C}_{ ext{Targe}}^t$	$_{\mathrm{t}}=\{y_{1},$	\ldots, y_{t-1}	$_{-1}\} \setminus \{y_t$	$\left\{ \cdot \right\}$.						
Copy UL	$\mathcal{L}_{CODVUL} =$	$-\Sigma_{t=1}^L$	$\Sigma_{c \in \mathcal{C}_{c}^{t}}$	$\log(1 - 1)$	$-P_{copy}(c \mathbf{y}_{1:t})$	$(-1, \mathbf{x}, \boldsymbol{\theta}))$					
	$ \mathcal{L}_{\text{CopyUL}} = -\Sigma_{t=1}^{L} \Sigma_{c \in \mathcal{C}_{\text{Copy}}^{t}} \log \left(1 - P_{copy}(c \mathbf{y}_{1:t-1}, \mathbf{x}, \boldsymbol{\theta})\right) $ Negative candidate list is composed of ground truth context tokens from previous time steps										
	L C			_	C		possibly copi	—	_		
	$\{y_1,\ldots,y_n\}$				_		pooloog copi	Ca). Copy	(9i 9i)		
	1		0 ± 10	0	$P(y_{t+k} \mathbf{y}_{1:t-1}$	<i>,</i>					
		<u>^</u>	1.	. •		1	and of time	$D_{\alpha\alpha\alpha}$			
Ahead	l'Io plan tl	ne surfa	ce realiz	zation of	f the output	sequence and	ead of time. 7	$V_k = \text{Decay } \mathbb{C}$	oemcient.		
	-				-	-	ead of time. 7	$V_k = \text{Decay } \mathbb{C}$	oemcient.		
Overall	$\mathcal{L} = \mathcal{L}_{K-1}$	StepMLE -	$+ \lambda_T \mathcal{L}_K$	C-StepTarg	$_{ ext{getUL}} + \lambda_C \mathcal{L}_K$	-StepCopyUL	,				
	$\mathcal{L} = \mathcal{L}_{K-1}$	StepMLE -	$+ \lambda_T \mathcal{L}_K$	C-StepTarg	$_{ ext{getUL}} + \lambda_C \mathcal{L}_K$	-StepCopyUL	rough the K-s				
Overall	$\mathcal{L} = \mathcal{L}_{K-1}$ Additiona	StepMLE -	$+ \lambda_T \mathcal{L}_K$ alize the	C-StepTarg	$_{ ext{getUL}} + \lambda_C \mathcal{L}_K$	–StepCopyUL epetitions th	,	step ahead U			
Overall	$\mathcal{L} = \mathcal{L}_{K-1}$ Additiona	StepMLE ⁻ Jlly, pend A ty Eval	$+ \lambda_T \mathcal{L}_K$ alize the	C–StepTarg e model	$_{ ext{getUL}} + \lambda_C \mathcal{L}_K$	–StepCopyUL epetitions th	rough the K-s	step ahead U			
Overall	$\mathcal{L} = \mathcal{L}_{K-1}$ Additional Quali	StepMLE ⁻ Jlly, pend A ty Eval	$+ \lambda_T \mathcal{L}_K$ alize the uation	C–StepTarg e model	$_{\text{getUL}} + \lambda_C \mathcal{L}_K$ for <i>future re</i>	-StepCopyUL epetitions th Divers	rough the K-s ity Evaluation	step ahead UI	L losses.		
Overall	$\mathcal{L} = \mathcal{L}_{K-1}$ Additiona Quali $P@M$	StepMLE ⁻ Jlly, pend A ty Eval	$+ \lambda_T \mathcal{L}_K$ alize the uation	C–StepTarg e model	$getUL + \lambda_C \mathcal{L}_K$ for <i>future re</i> %Duplicate	-StepCopyUL epetitions th Divers: %Duplicate	rough the K-s ity Evaluation	step ahead UI	L losses.		
Overall Loss	$\mathcal{L} = \mathcal{L}_{K-1}$ Additiona Quali $P@M$	StepMLE ⁻ Jlly, pend A ty Eval	$+ \lambda_T \mathcal{L}_K$ alize the uation	C—StepTarg e model #KPs	$getUL + \lambda_C \mathcal{L}_K$ for <i>future re</i> %Duplicate KPs ↓	-StepCopyUL epetitions th Divers %Duplicate Tokens↓	rough the K-s ity Evaluation Self-BLEU↓	step ahead U n Edit-Dist↓	L losses. Emb-Sim↓		
Overall Loss Ground Trutl	$\begin{array}{c} \mathcal{L} = \mathcal{L}_{K-1} \\ \text{Additiona} \\ \hline \mathbf{Quali} \\ P@M \\ \hline \mathbf{h} & - \end{array}$	StepMLE - Jlly, pend ty Eval R@M	$+ \lambda_T \mathcal{L}_K$ alize the uation $F_1 @ M$	C−StepTarg e model #KPs →5.3	$getUL + \lambda_C \mathcal{L}_K$ for <i>future re</i> %Duplicate KPs \downarrow 0.1	I-StepCopyUL Epetitions th Divers %Duplicate Tokens↓ 7.3	rough the K-s ity Evaluation Self-BLEU↓ 3.8	step ahead Ul n Edit-Dist↓ 32.7	L losses. Emb-Sim \downarrow 0.159		
Overall Loss Ground Trutl catSeq catSeqD catSeqCorr	$\mathcal{L} = \mathcal{L}_{K-1}$ Additiona $\boxed{\mathbf{Quali}}$ $P@M$ $h \qquad -$ 0.291	StepMLE - Jlly, pend ty Eval <i>R</i> @ <i>M</i> - 0.260	$+ \lambda_T \mathcal{L}_K$ alize the uation $F_1 @ M$ - 0.274	C−StepTarg e model #KPs →5.3 7.3	$getUL + \lambda_C \mathcal{L}_K$ for <i>future re</i> %Duplicate KPs ↓ 0.1 26.6	I-StepCopyUL epetitions th Diversa %Duplicate Tokens↓ 7.3 36.0	rough the K-s ity Evaluation Self-BLEU \downarrow 3.8 26.6	step ahead U \mathbf{n} Edit-Dist \downarrow 32.7 45.6	L losses. Emb-Sim \downarrow 0.159 0.328 0.325 0.309		
Overall Loss Ground Trutl catSeq catSeqD catSeqCorr catSeqTG	$\mathcal{L} = \mathcal{L}_{K-1}$ Additional Quali $P@M$ h - 0.291 0.294 0.283 0.295	StepMLE $-$ R@M - 0.260 0.257 0.264 0.262	$+ \lambda_T \mathcal{L}_K$ alize the uation $F_1 @ M$ - 0.274 0.274 0.274 0.273 0.278	C-StepTarge model #KPs $\rightarrow 5.3$ 7.3 6.7 7.0 6.8	getUL + $\lambda_C \mathcal{L}_K$ for future re %Duplicate KPs ↓ 0.1 26.6 25.7 23.2 24.7		rough the K-s ity Evaluation Self-BLEU \downarrow 3.8 26.6 27.0 24.5 26.2	step ahead UI n Edit-Dist \downarrow 32.7 45.6 45.3 44.0 45.2	L losses. Emb-Sim \downarrow 0.159 0.328 0.325 0.309 0.323		
Overall Loss Ground Trutl catSeq catSeqD catSeqCorr	$\mathcal{L} = \mathcal{L}_{K-1}$ Additional Quali $P@M$ h - 0.291 0.294 0.283 0.295	StepMLE - Jlly, pend ty Eval <i>R</i> @ <i>M</i> - 0.260 0.257 0.264	$+ \lambda_T \mathcal{L}_K$ alize the uation $F_1 @ M$ - 0.274 0.274 0.274 0.273	C-StepTarg e model #KPs $\rightarrow 5.3$ 7.3 6.7 7.0	getUL + $\lambda_C \mathcal{L}_K$ for future re %Duplicate KPs ↓ 0.1 26.6 25.7 23.2	Substant StepCopyUL StepCopyUL Substant Step $Point S = 10^{-10} \text{CopyUL}$ $Point S = 10^{-10} \text{CopyU}$ $Point S = 10^{-10} \text{CopyU}$	rough the K-s ity Evaluation Self-BLEU \downarrow 3.8 26.6 27.0 24.5	step ahead U n Edit-Dist \downarrow 32.7 45.6 45.3 44.0	L losses. Emb-Sim \downarrow 0.159 0.328 0.325 0.309		
Overall Loss Ground Trutl catSeq catSeqD catSeqCorr catSeqTG	$\mathcal{L} = \mathcal{L}_{K-1}$ Additional Quali $P@M$ h - 0.291 0.294 0.283 0.295 RF1 0.274	StepMLE $-$ R@M - 0.260 0.257 0.264 0.262	$+ \lambda_T \mathcal{L}_K$ alize the uation $F_1 @ M$ - 0.274 0.274 0.274 0.273 0.278	C-StepTarge model #KPs $\rightarrow 5.3$ 7.3 6.7 7.0 6.8	getUL + $\lambda_C \mathcal{L}_K$ for future re %Duplicate KPs ↓ 0.1 26.6 25.7 23.2 24.7		rough the K-s ity Evaluation Self-BLEU \downarrow 3.8 26.6 27.0 24.5 26.2	step ahead UI n Edit-Dist \downarrow 32.7 45.6 45.3 44.0 45.2	L losses. Emb-Sim \downarrow 0.159 0.328 0.325 0.309 0.323		
Overall Loss Ground Trut catSeq catSeqD catSeqCorr catSeqTG catSeqTG2R	$ \begin{array}{c} \mathcal{L} = \mathcal{L}_{K-1} \\ \text{Additiona} \\ \hline \mathbf{Quali} \\ P@M \\ \hline \mathbf{h} & - \\ 0.291 \\ 0.294 \\ 0.283 \\ 0.295 \\ \mathbf{C}F1 \\ 0.274 \\ \hline \mathbf{L} \\ 0.277 \\ \end{array} $	StepMLE - Jlly, pend ty Eval <i>R</i> @ <i>M</i> - 0.260 0.257 0.264 0.262 0.262 0.286	+ $\lambda_T \mathcal{L}_K$ alize the uation $F_1 @ M$ - 0.274 0.274 0.273 0.278 0.278 0.280	C-StepTarge model #KPs $\rightarrow 5.3$ 7.3 6.7 7.0 6.8 7.5	getUL + $\lambda_C \mathcal{L}_K$ for future re %Duplicate KPs ↓ 0.1 26.6 25.7 23.2 24.7 30.9		rough the K-s ity Evaluation Self-BLEU \downarrow 3.8 26.6 27.0 24.5 26.2 30.7	step ahead U \mathbf{n} Edit-Dist \downarrow 32.7 45.6 45.3 44.0 45.2 46.7	L losses. Emb-Sim \downarrow 0.159 0.328 0.325 0.309 0.323 0.341		

Table 2:KP generation results on KP20K dataset, evaluated on both quality and diversity criteria.

Ground Truth Target



Extensive experiments on datasets from 3 different domains demonstrate the effectiveness of our model for diverse keyphrase generation.

[1] Sean Welleck, Ilia Kulikov, Stephen Roller, Emily Dinan, Kyunghyun Cho, and Jason Weston. Neural text generation with unlikelihood training. In *ICLR*, 2020.

Email : hareeshbahuleyan@gmail.com Website : borealisai.com/en/research/ Source Code : tinyurl.com/divkgen



References

Contact Information





