Diverse Keyphrase Generation with Neural Unlikelihood Training

Hareesh Bahuleyan and Layla El Asri

Borealis AI, Montreal, Canada

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 → 26% keyphrase level duplication
- Existing solutions → 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

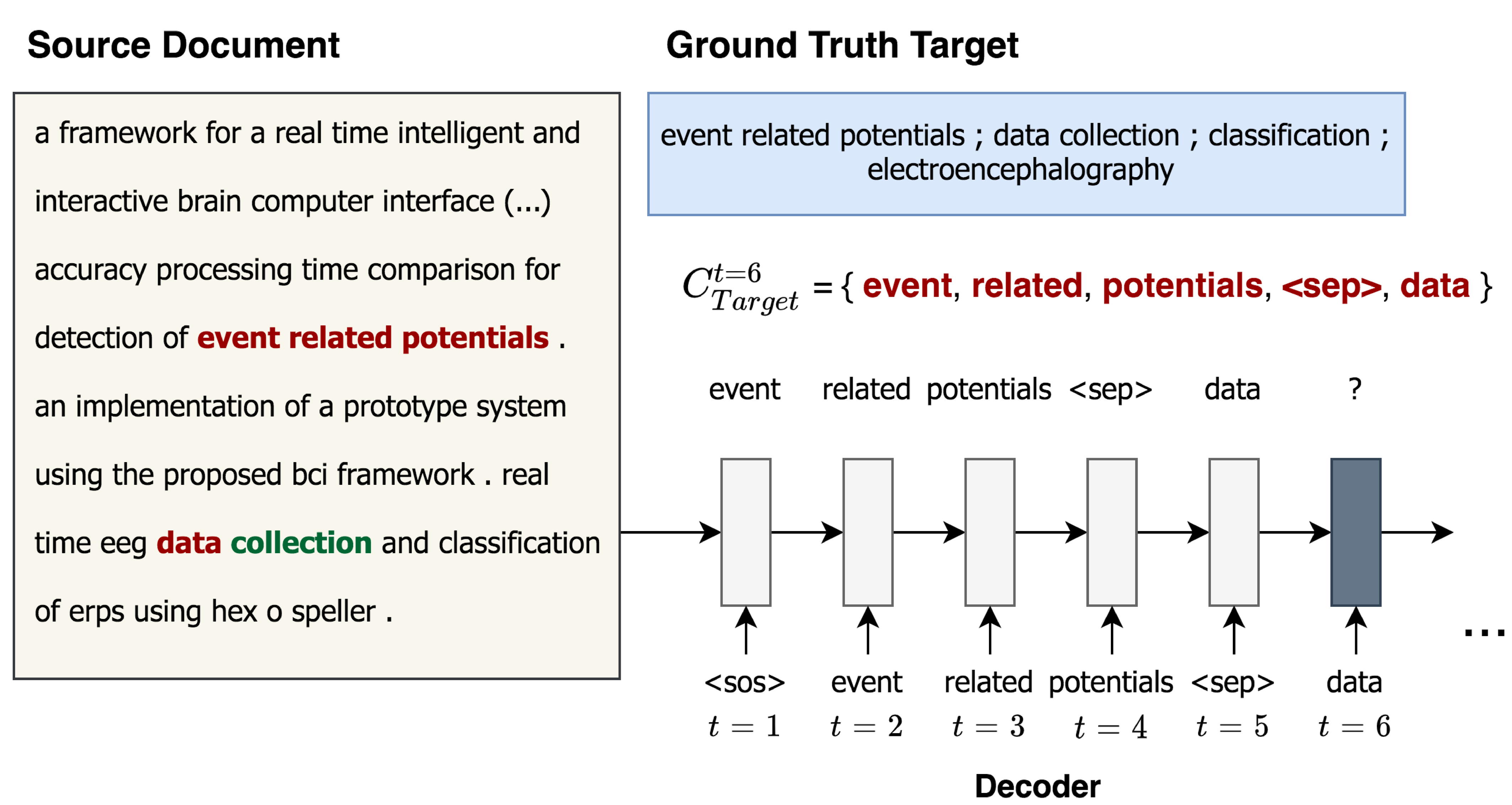
Diversity Evaluation Metrics

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

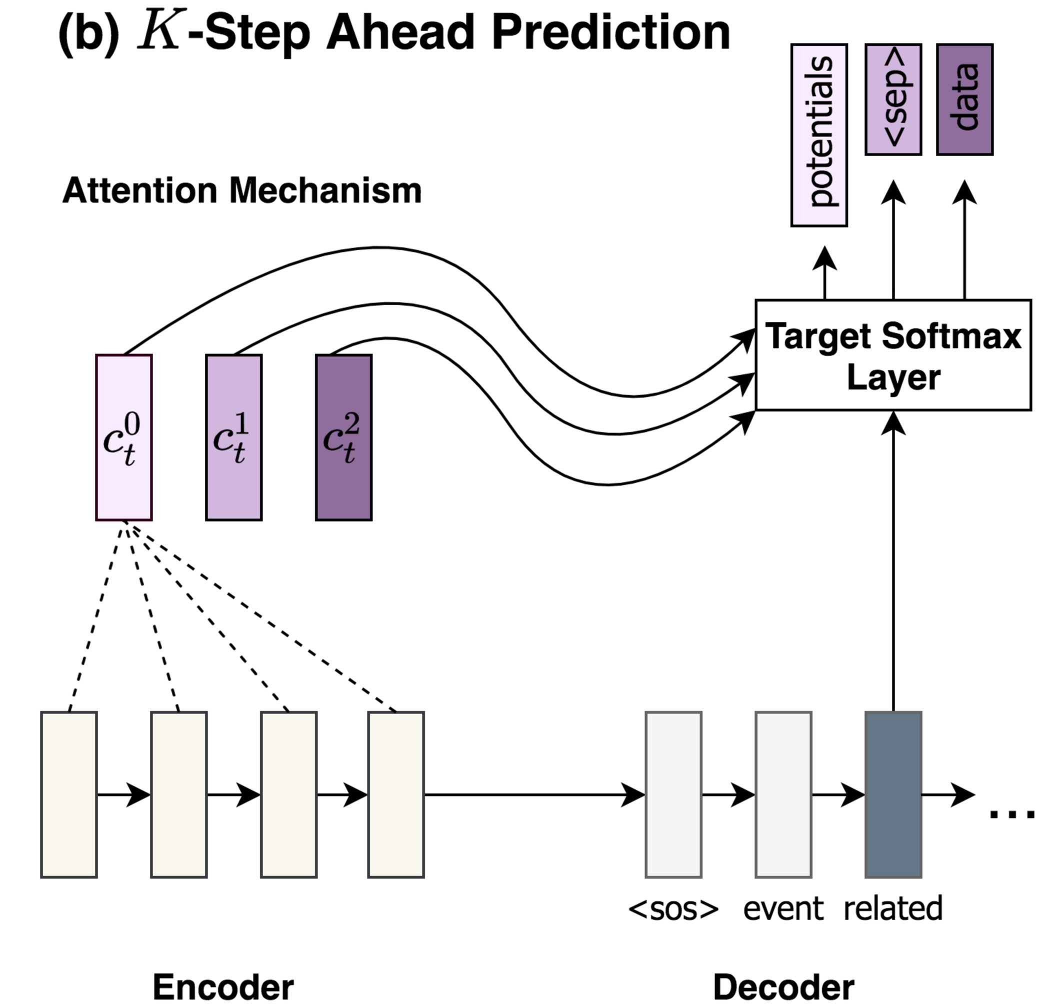
Ground Truth	*image segmentation ; region merging ; dynamic programming ; wald sequential probability ratio test*
catSeq MLE Baseline	*image segmentation ; region merging ; region merging ; dynamic programming ; image segmentation*
catSeqTG-2RF1 (RL)	*image segmentation ; region merging ; dynamic programming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

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

(a) Target + Copy Unlikelihood Training



(b) K-Step Ahead Prediction

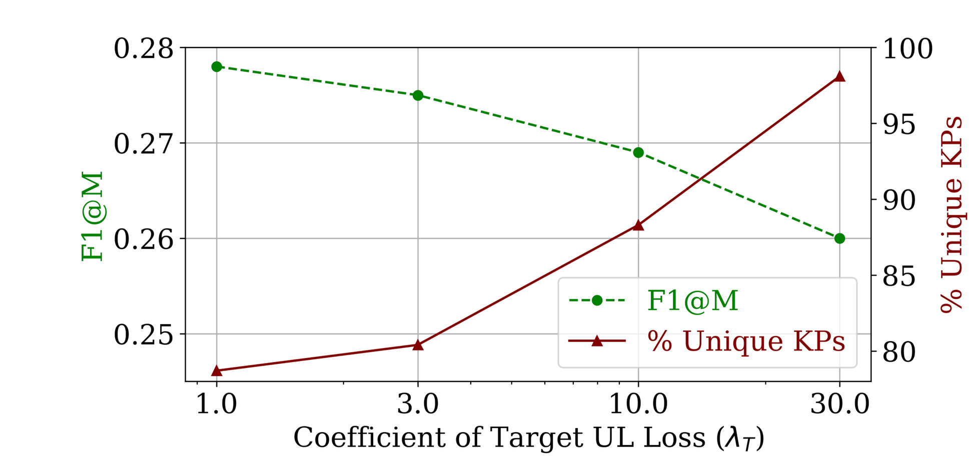


MLE	$\mathcal{L}_{\text{MLE}} = -\Sigma_{t=1}^{L} \log P(y_t \vert \mathbf{y}_{1:t-1}, \mathbf{x}, \boldsymbol{\theta})$ Next token prediction objective given input and context; $\boldsymbol{\theta}$ corresponds to the model parameters.
Target UL	$\mathcal{L}_{\text{TargetUL}} = -\Sigma_{t=1}^{L} \Sigma_{c \in \mathcal{C}^t_{\text{Target}}} \log \left(1 - P_{target}(c \vert \mathbf{y}_{1:t-1}, \mathbf{x}, \boldsymbol{\theta})\right)$ Negative candidate list consists of the ground truth context tokens from the previous time steps, i.e., $\mathcal{C}^t_{\text{Target}} = \{y_1, \dots, y_{t-1}\} \setminus \{y_t\}$.
Copy UL	$\mathcal{L}_{\text{CopyUL}} = -\Sigma_{t=1}^{L} \Sigma_{c \in \mathcal{C}^t_{\text{Copy}}} \log \left(1 - P_{copy}(c \vert \mathbf{y}_{1:t-1}, \mathbf{x}, \boldsymbol{\theta})\right)$ Negative candidate list is composed of ground truth context tokens from previous time steps *that also appear in the source text (and thus can be possibly copied).* $\mathcal{C}^t_{\text{Copy}} = \{y_i \mid y_i \in \{y_1, \dots, y_{t-1}\} \setminus \{y_t\}$ and $y_i \in \mathcal{V}_{\mathbf{x}}\}$
K-Step Ahead	$\mathcal{L}_{K\text{-StepMLE}} = -\Sigma_{t=1}^{L} \Sigma_{k=0}^{K} \gamma_k \log P(y_{t+k} \vert \mathbf{y}_{1:t-1}, \mathbf{x}, \boldsymbol{\theta})$ To plan the surface realization of the output sequence ahead of time. γ_k = Decay Coefficient.
Overall Loss	$\mathcal{L} = \mathcal{L}_{K\text{-StepMLE}} + \lambda_T \mathcal{L}_{K\text{-StepTargetUL}} + \lambda_C \mathcal{L}_{K\text{-StepCopyUL}}$ Additionally, penalize the model for *future repetitions* through the K-step ahead UL losses.

	Quality Evaluation			Diversity Evaluation					
	$P@M$	$R@M$	$F_1@M$	#KPs	%Duplicate KPs ↓	%Duplicate Tokens ↓	Self-BLEU ↓	Edit-Dist ↓	Emb-Sim ↓
Ground Truth	-	-	-	→5.3	0.1	7.3	3.8	32.7	0.159
catSeq	0.291	0.260	0.274	7.3	26.6	36.0	26.6	45.6	0.328
catSeqD	0.294	0.257	0.274	6.7	25.7	35.3	27.0	45.3	0.325
catSeqCorr	0.283	0.264	0.273	7.0	23.2	33.5	24.5	44.0	0.309
catSeqTG	**0.295**	0.262	0.278	6.8	24.7	34.3	26.2	45.2	0.323
catSeqTG-2RF1	0.274	**0.286**	**0.280**	7.5	30.9	41.7	30.7	46.7	0.341
DivKGen (UL)	0.277	0.261	0.269	**5.0**	5.3	12.6	9.7	**34.4**	**0.181**
+K-StepMLE	0.274	0.239	0.255	4.6	6.1	13.9	11.5	36.2	0.197
+K-StepUL	0.273	0.240	0.256	4.6	**4.9**	**11.7**	**8.8**	35.2	0.185

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

Quality-Diversity Trade-off



Conclusions

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

References

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

Contact Information

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

Code URL

BOREALIS AI

RBC