Quality-Aware Decoding for Neural Machine Translation

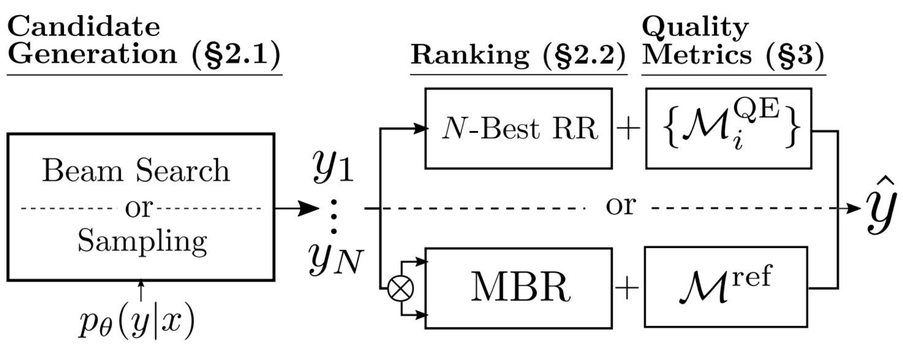
Carnegie Mellon University

Patrick Fernandes*, António Farinhas*, Ricardo Rei,
José G. C. de Souza, Perez Ogayo, Graham Neubig,
André F. T. Martins

TÉCNICO LISBOA

Tired of beam search?

We explore an alternative decoding method that leverages neural metrics to produce better translations!



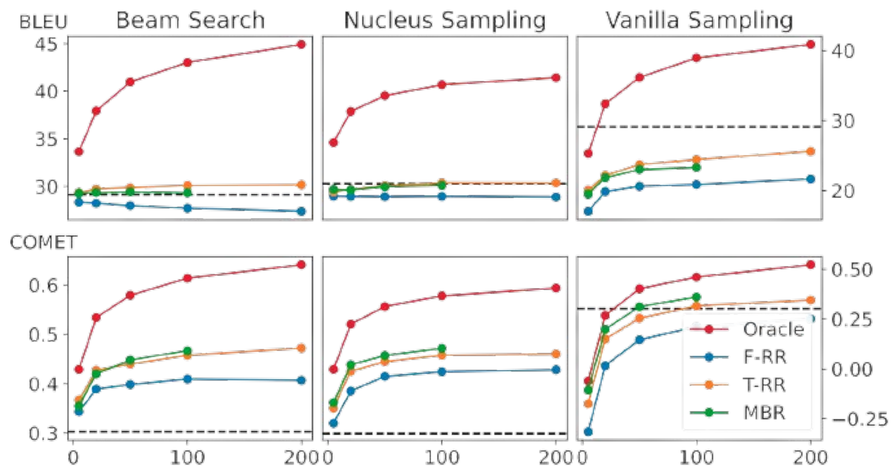

```bash
# generate candidates
fairseq-generate ... --nbest $nbest \
    | grep ^H | cut -c 3- | sort -n | cut -f3- \
    > $cands

# apply mbr
qaware-mbr $cands --src $src -n $nbest --metric comet \
    > $translations
```

Quality-aware decoding framework

1) Generate translation candidates according to the model;

2) Use reference-free and/or reference-based MT metrics to rank these candidates;

3) Pick the highest ranked candidate as the final translation.



Automatic evaluation metrics

	Large (WMT20)				Small (IWSLT)			
	BLEU	chrF	BLEURT	COMET	BLEU	chrF	BLEURT	COMET
Baseline	**36.01**	63.88	0.7376	0.5795	29.12	56.23	0.6635	0.3028
F-RR w/ COMET-QE	29.83	59.91	0.7457	0.6012	27.38	54.89	0.6848	0.4071
F-RR w/ MBART-QE	32.92	62.71	0.7384	0.5831	27.30	55.62	0.6765	0.3533
F-RR w/ OpenKiwi	30.38	59.56	0.7401	0.5623	25.35	51.53	0.6524	0.2200
F-RR w/ Transquest	31.28	60.94	0.7368	0.5739	26.90	54.46	0.6613	0.2999
T-RR w/ BLEU	35.34	63.82	0.7407	0.5891	**30.51**	**57.73**	0.7077	0.4536
T-RR w/ BLEURT	33.39	62.56	0.7552	0.6217	30.16	57.40	0.7127	0.4741
T-RR w/ COMET	34.26	63.31	0.7546	0.6276	30.16	57.32	0.7124	0.4721
MBR w/ BLEU	34.94	63.21	0.7333	0.5680	29.25	56.36	0.6619	0.3017
MBR w/ BLEURT	32.90	62.34	0.7649	0.6047	28.69	56.28	0.7051	0.3799
MBR w/ COMET	33.04	62.65	0.7477	0.6359	29.43	56.74	0.6882	0.4480
T-RR+MBR w/ BLEU	35.84	**63.96**	0.7395	0.5888	30.23	57.34	0.6913	0.3969
T-RR+MBR w/ BLEURT	33.61	62.95	**0.7658**	0.6165	29.28	56.77	**0.7225**	0.4361
T-RR+MBR w/ COMET	34.20	63.35	0.7526	**0.6418**	29.46	57.13	0.7058	**0.5005**

Human evaluation

	EN-DE (WMT20)				EN-RU (WMT20)			
	Minor	Major	Critical	MQM	Minor	Major	Critical	MQM
Reference	24	67	0	97.04	5	11	0	99.30
Baseline	8	139	0	95.66	17	239	49	79.78
F-RR w/ COMET-QE	15	204	0	93.47	13	254	80	76.25
T-RR w/ COMET	12	109	0	**96.20**	9	141	45	85.97†
MBR w/ COMET	11	161	0	94.38	8	182	40	83.65
T-RR + MBR w/ COMET	10	138	0	95.44	11	134	45	**86.78**†

Acknowledgments:
P2020 program MAIA (LISBOA-01-0247- FEDER-045909), European Research Council (ERC StG DeepSPIN 758969), European Union's Horizon 2020 research and innovation program (QUARTZ grant agreement 951847), and Fundação para a Ciência e Tecnologia through UIDB/50008/2020.