WMT 2023 - Biomedical task

September 5, 2023

1 Results for Automatic Evaluation of the MEDLINE test sets

We separate results into two sections: (i) results for submissions to the WMT submission system; (ii) results for submissions in OCELoT.

BLEU scores were calculated using the multi-eval tool and tokenization as provided in Moses. In all result tables, an asterisk * indicates the primary run, as informed by the participants, in the case of multiple runs.

1.1 WMT Submission system

Teams	Runs	en2de	en2fr
HuaweiTSC	run1	0.4154	
HuaweiTSC	run2	0.4127	
HuaweiTSC	run3	0.4198	
ustc_ml_group	run1		0.4869
ustc_ml_group	run2		0.4974

Table 1: BLEU scores for test sentences, from English.

Teams	Runs	de2en	fr2en
HuaweiTSC	run1	0.4266	
HuaweiTSC	run2	0.4266	
HuaweiTSC	run3	0.4338	
NRPU_FJWU	run1		0.3148
NRPU_FJWU(1)	run1		0.2874
NRPU_FJWU	run2		0.2050
NRPU_FJWU	run3		0.2246
NRPU_FJWU(1)	run3		0.3148
ustc_ml_group	run1		0.3889
ustc_ml_group	run2		0.3620

Table 2: BLEU scores for test sentences, to English.

Teams	Runs	en2de	en2fr
HuaweiTSC	run1	0.4369	
HuaweiTSC	run2	0.4345	
HuaweiTSC	run3	0.4422	
ustc_ml_group	run1		0.4908
ustc_ml_group	run2		0.4998

Table 3: BLEU scores for "OK" aligned test sentences, from English.

Teams	Runs	de2en	fr2en
HuaweiTSC	run1	0.4778	
HuaweiTSC	run2	0.4776	
HuaweiTSC	run3	0.4853	
NRPU_FJWU	run1		0.3350
NRPU_FJWU(1)	run1		0.3082
NRPU_FJWU	run2		0.2202
NRPU_FJWU	run3		0.2395
NRPU_FJWU(1)	run3		0.3350
ustc_ml_group	run1		0.4124
ustc_ml_group	run2		0.3854

Table 4: BLEU scores for "OK" aligned test sentences, to English.

1.2 OCELoT submissions

Teams	Runs	en2de	en2es	en2fr	en2it	en2pt	en2ru
AIRC		0.3443					
GPT4-5shot		0.3881					0.3649
Lan-BridgeMT		0.3463	0.5098	0.5164	0.4640	0.4832	0.3361
NLLB_Greedy		0.3663					0.3461
NLLB_MBR_BLEU		0.3625					0.3504
ONLINE-A		0.4332					0.4125
ONLINE-B		0.4298					0.4648
ONLINE-G		0.4263					0.3939
ONLINE-M		0.3984					0.3827
ONLINE-W		0.4451					0.4083
ONLINE-Y		0.4075					0.4049
PROMT							0.3872
UPCite-CLILLF				0.2706			
ZengHuiMT		0.3883					0.3775

Table 5: BLEU scores for "OK" aligned test sentences, from English.

Teams	Runs	de2en	es2en	fr2en	it2en	pt2en	ru2en
AIRC		0.3714					
GPT4-5shot		0.4371					0.4774
GTCOM_Peter		0.4212					
HW-TSC		0.4771					
Lan-BridgeMT		0.4215	0.5769	0.4323	0.5272	0.5569	0.4750
NLLB_Greedy		0.4040					0.4386
NLLB_MBR_BLEU		0.3992					0.4437
ONLINE-A		0.4606					0.5723
ONLINE-B		0.4662					0.4648
ONLINE-G		0.4364					0.5445
ONLINE-M		0.4465					0.4607
ONLINE-W		0.4759					0.4919
ONLINE-Y		0.4075					0.5089
PROMT							0.5156
UPCite-CLILLF				0.0093			
ZengHuiMT		0.4316					0.5256

Table 6: BLEU scores for "OK" aligned test sentences, to English.