

A. Benchmarks for downstream evaluation

A summary of benchmarks for downstream evaluation is shown in Table 3.

B. Supplementary results

B.1. Per-language results of cross-lingual evaluation

Table 4 shows the overall performance of cross-lingual evaluation using LMs and MT models. Note that models are fine-tuned only in English and evaluated in other languages; moreover benchmarks differ in which languages they include. As a consequence, some scores are not available for some languages.

B.2. The scaling effect of MT as continued training in different layers

Table 5 shows the results of the first 11 layers in the encoders of mBART series models. Similar to the analysis in Section 3.3, we calculate the norm of the vectorized diagonal matrices of singular values $\text{diag}(\Sigma)$ and their pairwise distance to the corresponding vectors derived from the same weight matrices in The transformer attention module and fully connected layers in the base mBART model. The results indicate the same conclusion drawn from the analysis on the 12th layer.

Task	# of Languages	Train ^{en}	Dev ^{avg}	Test ^{avg}	Metric	Data Source
NER	4	15.0K	2.8K	3.4K	F1	ECI Multilingual Text Corpus
POS	18	25.4K	1.0K	0.9K	ACC	UD Tree-banks (v2.5)
NC	5	100K	10K	10K	ACC	Commercial News Website
XNLI	15	433K	2.5K	5K	ACC	MultiNLI Corpus
PAWS-X	4	49.4K	2K	2K	ACC	Wikipedia
QADSM	3	100K	10K	10K	ACC	Commercial Search Engine
WPR	7	100K	10K	10K	nDCG	Commercial Search Engine
QAM	3	100K	10K	10K	ACC	Commercial Search Engine

Table 3: A summary of benchmarks for downstream evaluation. We choose 8 downstream tasks from XGLUE (Liang et al., 2020) for cross-lingual evaluation and x tasks for monolingual evaluation. The training set of each task is only available in English, with |Train|^{en} denoting the number of labeled instances. |Dev|^{avg} and |Test|^{avg} denote the average numbers of labeled instances in the dev sets and test sets, respectively.

Task	Model	AR	BG	DE	EL	EN	ES	FR	HI	IT	NL	PL	PT	RU	SW	TH	TR	UR	VI	ZH	AVG	
NC	mBERT	-	-	81.09	-	91.98	80.73	75.84	-	-	-	-	-	76.96	-	-	-	-	-	-	81.32	
	XLM-R	-	-	81.79	-	91.97	82.14	76.10	-	-	-	-	-	78.63	-	-	-	-	-	-	82.13	
	mBART	-	-	82.82	-	91.90	81.37	75.95	-	-	-	-	-	78.43	-	-	-	-	-	-	82.09	
	mBART m2o	-	-	80.32	-	91.46	78.95	74.11	-	-	-	-	-	76.96	-	-	-	-	-	-	80.36	
	mBART o2m	-	-	63.63	-	91.39	69.12	65.40	-	-	-	-	-	37.60	-	-	-	-	-	-	65.43	
	mBART m2m	-	-	77.44	-	91.78	75.02	71.99	-	-	-	-	-	75.19	-	-	-	-	-	-	78.28	
XNLI	NLLB 600M	-	-	67.52	-	91.79	76.43	71.79	-	-	-	-	-	72.41	-	-	-	-	-	-	75.99	
	mBERT	62.49	66.79	71.04	65.34	82.29	74.06	73.94	58.92	-	-	-	-	65.62	51.24	51.16	61.41	56.27	68.31	68.63	65.17	
	XLM-R	70.68	76.27	76.06	74.70	84.86	79.60	78.31	68.15	-	-	-	-	74.18	63.78	71.29	71.89	65.34	74.10	73.21	73.49	
	mBART	69.40	56.95	75.14	34.82	84.22	78.59	75.82	66.14	-	-	-	-	73.86	58.03	67.31	68.80	62.25	71.65	71.49	67.63	
	mBART m2o	68.96	58.76	74.90	36.87	82.13	76.59	75.62	65.78	-	-	-	-	73.09	42.33	59.72	68.39	60.80	71.16	73.01	65.87	
	mBART o2m	42.01	43.49	49.28	34.86	81.85	53.78	55.54	40.00	-	-	-	-	54.70	39.80	39.84	41.57	35.86	56.02	52.45	48.07	
PAWS-X	mBART m2m	58.39	50.72	68.71	35.34	83.90	66.31	73.78	65.02	-	-	-	-	57.51	40.84	49.52	60.32	58.71	70.88	62.53	60.17	
	NLLB 600M	68.47	69.88	63.45	72.53	81.12	72.29	72.73	68.76	-	-	-	-	72.37	58.35	64.94	59.28	64.30	68.76	67.63	68.32	
	mBERT	-	-	82.20	-	92.85	84.60	86.70	-	-	-	-	-	-	-	-	-	-	-	-	86.59	
	XLM-R	-	-	85.75	-	93.40	88.30	87.95	-	-	-	-	-	-	-	-	-	-	-	-	88.85	
	mBART	-	-	86.70	-	93.65	88.30	88.30	-	-	-	-	-	-	-	-	-	-	-	-	89.24	
	mBART m2o	-	-	81.80	-	91.00	84.50	85.20	-	-	-	-	-	-	-	-	-	-	-	-	85.63	
QAM	mBART o2m	-	-	76.35	-	89.90	78.90	81.45	-	-	-	-	-	-	-	-	-	-	-	-	81.65	
	mBART m2m	-	-	83.95	-	92.20	85.75	86.80	-	-	-	-	-	-	-	-	-	-	-	-	87.18	
	NLLB 600M	-	-	67.40	-	82.35	71.65	72.00	-	-	-	-	-	-	-	-	-	-	-	-	73.35	
	mBERT	-	-	62.08	-	69.47	-	62.35	-	-	-	-	-	-	-	-	-	-	-	-	64.63	
	XLM-R	-	-	66.98	-	69.69	-	65.45	-	-	-	-	-	-	-	-	-	-	-	-	67.37	
	mBART	-	-	66.36	-	70.46	-	66.71	-	-	-	-	-	-	-	-	-	-	-	-	67.84	
QADSM	mBART m2o	-	-	64.20	-	65.10	-	62.39	-	-	-	-	-	-	-	-	-	-	-	-	63.90	
	mBART o2m	-	-	55.27	-	65.41	-	54.60	-	-	-	-	-	-	-	-	-	-	-	-	58.43	
	mBART m2m	-	-	62.62	-	66.21	-	60.72	-	-	-	-	-	-	-	-	-	-	-	-	63.18	
	NLLB 600M	-	-	57.90	-	66.47	-	60.14	-	-	-	-	-	-	-	-	-	-	-	-	61.50	
	mBERT	-	-	59.94	-	67.04	-	62.30	-	-	-	-	-	-	-	-	-	-	-	-	63.09	
	XLM-R	-	-	63.19	-	71.44	-	66.02	-	-	-	-	-	-	-	-	-	-	-	-	66.88	
WPR	mBART	-	-	61.83	-	69.83	-	64.79	-	-	-	-	-	-	-	-	-	-	-	-	65.48	
	mBART m2o	-	-	63.15	-	64.07	-	64.34	-	-	-	-	-	-	-	-	-	-	-	-	63.85	
	mBART o2m	-	-	63.40	-	65.17	-	59.61	-	-	-	-	-	-	-	-	-	-	-	-	62.73	
	mBART m2m	-	-	60.89	-	65.45	-	62.05	-	-	-	-	-	-	-	-	-	-	-	-	62.80	
	NLLB 600M	-	-	64.48	-	64.45	-	62.71	-	-	-	-	-	-	-	-	-	-	-	-	63.88	
	mBERT	-	-	76.64	-	77.29	75.07	73.92	-	66.58	-	-	-	77.04	-	-	-	-	-	-	62.67	74.42
NER	XLM-R	-	-	77.08	-	77.79	76.14	74.94	-	67.87	-	-	-	77.93	-	-	-	-	-	-	62.81	75.29
	mBART	-	-	76.74	-	77.18	75.41	74.22	-	67.40	-	-	-	77.38	-	-	-	-	-	-	62.86	74.72
	mBART m2o	-	-	75.60	-	76.17	74.08	73.31	-	66.21	-	-	-	76.59	-	-	-	-	-	-	62.38	73.66
	mBART o2m	-	-	75.32	-	75.99	74.07	72.76	-	65.39	-	-	-	75.80	-	-	-	-	-	-	61.24	73.22
	mBART m2m	-	-	76.22	-	76.22	74.28	73.23	-	66.35	-	-	-	75.79	-	-	-	-	-	-	61.93	73.68
	NLLB 600M	-	-	76.01	-	76.35	73.81	73.48	-	65.84	-	-	-	76.46	-	-	-	-	-	-	62.02	73.66
POS	mBERT	-	-	68.84	-	90.78	73.27	-	-	-	77.28	-	-	-	-	-	-	-	-	-	77.54	
	XLM-R	-	-	69.99	-	90.45	75.77	-	-	-	78.62	-	-	-	-	-	-	-	-	-	78.71	
	mBART	-	-	71.31	-	91.35	72.55	-	-	-	75.57	-	-	-	-	-	-	-	-	-	77.70	
	mBART m2o	-	-	52.41	-	89.61	50.71	-	-	-	53.36	-	-	-	-	-	-	-	-	-	61.52	
	mBART o2m	-	-	25.66	-	89.22	53.31	-	-	-	52.13	-	-	-	-	-	-	-	-	-	55.08	
	mBART m2m	-	-	65.25	-	88.99	66.86	-	-	-	66.58	-	-	-	-	-	-	-	-	-	71.92	
POS	NLLB 600M	-	-	29.4	-	89.46	43.21	-	-	-	54.83	-	-	-	-	-	-	-	-	-	54.23	
	mBERT	57.26	85.84	90.21	82.61	95.84	87.67	85.80	66.57	91.78	87.78	80.93	88.93	80.57	-	41.88	68.87	60.12	55.09	60.19	76.00	
	XLM-R	69.44	88.70	91.75	87.63	96.43	88.20	89.22	72.10	91.35	88.46	83.82	90.07	87.12	-	58.08	72.76	64.28	57.06	58.45	79.72	
	mBART	63.55	71.76	90.56	29.74	96.13	87.07	87.75	67.61	90.64	87.51	80.60	88.29	83.35	-	55.56	66.53	55.61	54.62	51.56	72.69	
	mBART m2o	63.97	71.30	90.64	24.82	95.74	84.98	85.19	64.32	87.45	86.18	80.12	82.91	81.91	-	51.13	66.77	50.65	52.29	53.18	70.75	
	mBART o2m	53.78	58.78	61.97	41.62	95.63	63.08	70.43	48.16	59.92	60.01	53.62	58.11	61.60	-	37.12	46.21	42.90	44.95	44.08	55.67	
POS	mBART m2m	64.60	71.58	90.35	21.66	96.06	81.21	86.20	65.94	83.71	85.30	81.28	81.65	84.81	-	41.83	63.55	52.44	51.02	51.33	69.70	
	NLLB 600M	63.76	84.47	77.64	79.22	96.12	82.61	83.36	66.59	84.93	75.9	74.57	80.95	80.92	-	46.56	56.46	58.59	45.63	46.32	71.37	

Table 4: The overall performance of cross-lingual natural language understanding. We use the base architecture for mBERT and XLM-R. mBART models only utilize the 12-layer encoders. ‘m2o’ means many-to-one. ‘o2m’ means one-to-many. ‘m2m’ means many-to-many. ‘-’ denotes that the benchmark does not cover the corresponding language.

Layer	Model	K		Q		V		Out		FC1		FC2	
		$\ \sigma\ $	d	$\ \sigma\ $	d	$\ \sigma\ $	d	$\ \sigma\ $	d	$\ \sigma\ $	d	$\ \sigma\ $	d
0	mBART	39.25	0.00	41.16	0.00	32.82	0.00	30.30	0.00	88.19	0.00	66.57	0.00
	mBART m2m	48.46	9.55	48.83	7.71	31.01	2.14	29.29	3.10	93.43	6.64	73.70	8.30
	mBART m2o	48.49	9.59	48.95	7.90	30.51	2.44	28.99	3.55	93.94	7.12	74.39	9.03
	mBART o2m	49.15	10.62	50.54	10.50	35.90	3.87	35.75	6.73	111.30	23.86	94.52	28.87
1	mBART	44.41	0.00	46.89	0.00	29.21	0.00	31.80	0.00	90.87	0.00	69.86	0.00
	mBART m2m	51.21	7.58	52.62	5.79	27.69	1.73	30.56	2.41	96.44	6.34	76.99	8.09
	mBART m2o	51.64	8.07	53.10	6.36	27.49	1.78	30.47	2.63	97.06	7.31	77.83	9.07
	mBART o2m	54.00	9.77	56.38	9.70	35.17	6.40	37.90	6.64	113.14	23.36	96.06	26.69
2	mBART	46.56	0.00	47.80	0.00	31.80	0.00	32.22	0.00	93.76	0.00	71.51	0.00
	mBART m2m	51.83	5.86	52.66	5.04	31.34	1.32	31.96	2.48	99.62	6.49	78.89	8.19
	mBART m2o	52.40	6.53	53.22	5.66	31.26	1.09	31.94	2.53	100.24	7.50	79.71	9.18
	mBART o2m	56.22	9.75	57.47	9.91	38.20	6.91	39.20	7.45	114.89	21.77	96.37	25.03
3	mBART	48.09	0.00	48.65	0.00	37.77	0.00	36.85	0.00	96.27	0.00	73.45	0.00
	mBART m2m	52.57	4.54	53.05	4.63	39.06	1.65	38.06	2.42	101.80	6.09	80.15	7.27
	mBART m2o	53.15	5.15	53.62	5.21	39.15	1.59	38.12	2.40	102.31	7.18	80.84	8.12
	mBART o2m	56.82	8.80	57.69	9.28	45.90	8.25	45.43	9.00	117.17	21.21	97.57	24.24
4	mBART	51.59	0.00	51.06	0.00	40.80	0.00	38.91	0.00	99.22	0.00	77.23	0.00
	mBART m2m	56.45	5.05	55.97	5.08	43.02	2.51	41.20	2.71	104.97	6.76	84.00	7.36
	mBART m2o	57.10	5.70	56.62	5.72	43.00	2.31	41.17	2.50	105.74	8.23	85.01	8.72
	mBART o2m	61.88	10.40	61.58	10.61	49.68	8.98	48.54	9.82	121.06	21.94	101.56	24.42
5	mBART	59.93	0.00	58.27	0.00	42.09	0.00	38.29	0.00	101.79	0.00	84.25	0.00
	mBART m2m	65.19	5.44	63.76	5.69	44.03	2.11	40.23	2.61	107.87	7.42	91.05	7.56
	mBART m2o	65.86	6.14	64.44	6.33	44.22	2.21	40.42	2.63	108.94	8.86	92.49	9.56
	mBART o2m	70.59	10.94	69.24	11.26	51.39	9.36	48.64	10.70	124.78	23.18	108.30	24.15
6	mBART	57.55	0.00	55.80	0.00	45.27	0.00	40.54	0.00	101.52	0.00	89.89	0.00
	mBART m2m	61.71	4.60	60.09	4.47	48.11	2.94	43.59	3.34	108.45	7.71	97.85	8.11
	mBART m2o	62.22	5.12	60.62	4.93	48.44	3.24	43.95	3.63	109.45	8.79	99.23	9.60
	mBART o2m	67.76	10.73	66.31	10.74	55.16	9.96	51.59	11.24	125.45	24.20	113.90	24.10
7	mBART	53.94	0.00	52.44	0.00	49.99	0.00	46.20	0.00	98.28	0.00	96.88	0.00
	mBART m2m	57.72	4.33	56.30	4.04	52.77	2.98	49.09	3.36	106.68	8.96	105.70	8.94
	mBART m2o	58.03	4.61	56.64	4.31	53.23	3.32	49.58	3.72	107.94	10.21	107.26	10.47
	mBART o2m	64.09	10.79	62.87	10.64	59.23	9.49	56.08	10.17	123.84	25.99	120.47	23.76
8	mBART	48.50	0.00	47.79	0.00	51.22	0.00	49.37	0.00	94.92	0.00	100.91	0.00
	mBART m2m	51.54	3.69	50.86	3.51	54.22	3.17	52.37	3.14	104.18	9.91	109.97	9.35
	mBART m2o	51.67	3.75	51.00	3.60	54.76	3.63	52.96	3.67	105.67	11.31	111.65	10.98
	mBART o2m	58.53	10.57	57.90	10.39	60.62	9.65	58.81	9.60	121.89	27.64	124.33	23.66
9	mBART	44.39	0.00	44.78	0.00	50.78	0.00	49.83	0.00	94.01	0.00	103.16	0.00
	mBART m2m	47.47	3.73	47.82	3.55	53.04	2.78	52.13	2.95	103.53	10.31	112.32	9.64
	mBART m2o	47.57	3.76	47.92	3.56	53.55	3.13	52.61	3.15	105.10	11.78	113.97	11.19
	mBART o2m	55.12	11.12	55.55	11.12	58.83	8.49	57.86	8.42	120.99	27.80	126.63	23.92
10	mBART	43.50	0.00	42.87	0.00	55.41	0.00	54.97	0.00	91.92	0.00	105.62	0.00
	mBART m2m	46.93	4.66	46.28	4.31	57.64	2.89	57.30	4.13	101.92	10.67	114.82	9.94
	mBART m2o	46.88	4.50	46.24	4.23	58.24	3.25	57.93	4.21	103.56	12.19	116.47	11.45
	mBART o2m	54.68	12.00	54.22	11.91	62.95	8.24	62.28	8.64	119.35	28.09	128.21	23.29

Table 5: The scaling effect via singular value decomposition of mBART and the continued-trained MT models.