

A Performance Evaluation reporting on Precision, Recall and F1-Score

Tables 1, 2 and 3 show a fine-grained analysis for English, Chinese and Arabic respectively.

Feature Type	Model	Train Lang	Embed -dings	Trigger Identification (%)			Trigger Classification(%)		
				P	R	F1	P	R	F1
Disc. Only	<i>Ji's Cross-entity'08</i>	EN	-	N/A			72.9	64.3	68.3
	<i>Liao's Cross-event'10</i>			N/A			68.7	68.9	68.8
	<i>Li's Joint-Beam'13</i>			76.9	65.0	70.4	73.7	62.3	67.5
Disc. + Cont.	<i>Chen's DMCNN'15</i>		Skip-Gram	80.4	67.7	73.5	75.6	63.6	69.1
	<i>Nguyen's JRNN'16</i>		C-BOW	68.5	75.7	71.9	66.0	73.0	69.3
	<i>Lius's JMEE'18</i>		Glove	80.2	72.1	75.9	76.3	71.3	73.7
	<i>Zhang's GAIL'19</i>		ELMo	76.8	71.2	73.9	74.8	69.4	72.0
Cont. Only	<i>Feng's HNN'16</i>		Skip-Gram	80.8	71.5	75.9	84.6	64.9	73.4
	<i>Liu's GMLATT'18</i>		Skip-Gram	80.9	68.1	74.1	78.9	66.9	72.4
Our Method	Bi-LSTM-Char-CRF		EN	FastText	67.7	67.2	67.5	63.4	63.0
		EN	MUSE	69.4	68.4	68.9	62.9	62.0	62.5
		EN+ZH		77.8	62.7	69.5	73.7	59.4	65.8
		EN+AR		74.1	67.5	70.6	70.2	63.9	66.9
		All		63.1	70.3	66.5	58.5	65.1	61.6
	EN	Base	77.6	80.8	79.2	73.8	76.9	75.3	
	BERT-CRF	EN	BERT (multi)	79.5	76.2	77.8	74.8	71.6	73.1
		EN+ZH		78.8	80.9	79.8	74.3	76.1	75.2
		EN+AR		81.4	77.4	79.3	76.4	72.6	74.5
		All		85.6	73.8	79.2	79.7	68.3	73.5

Table 1: Comparison of performance **testing on English** using prior work baselines in the first half and our method using Bi-LSTM-Char-CRF with MUSE embeddings, BERT-CRF in the 2nd half.

Feature Type	Model	Train Lang	Embed -dings	Trigger Identification (%)			Trigger Classification(%)			
				P	R	F1	P	R	F1	
Disc. Only	<i>Li's MaxEnt'13</i>	ZH	-	50.0	77.0	60.6	47.5	73.1	57.6	
	<i>Chen's Rich-C'12</i>			62.2	71.9	66.7	58.9	68.1	63.2	
Disc. + Cont.	<i>Hsi's multi'16</i>		multi_proj	N/A			44.3	20.9	39.6	
Cont. Only	<i>Feng's HNN'16</i>		Skip-Gram	74.2	63.1	68.2	77.1	53.1	63.0	
Our Method	Bi-LSTM-Char-CRF		ZH	FastText	89.7	83.8	86.6	68.6	64.5	69.5
			ZH	MUSE	28.6	30.8	29.6	24.1	25.9	25.0
			EN		71.5	53.6	61.3	56.9	42.7	48.8
			EN+ZH		83.9	71.5	77.2	76.7	65.4	70.6
			All		71.3	73.9	72.6	63.2	65.5	64.3
			ZH		Base	76.4	94.3	84.4	72.3	89.2
		BERT-CRF	ZH	BERT (multi)	76.4	92.6	83.7	72.8	88.2	79.8
			EN		66.5	90.9	76.8	59.3	81.0	68.5
			EN+ZH		76.4	94.9	84.7	73.3	91.1	81.2
			All		80.9	95.7	87.7	76.8	90.9	83.2

Table 2: Comparison of performance **testing on Chinese** using prior work baselines in the first half and our method using Bi-LSTM-Char-CRF with MUSE embeddings, BERT-CRF in the second half.

Model	Train Lang	Embed -dings	Trigger Identification (%)			Trigger Classification(%)		
			P	R	F1	P	R	F1
Bi-LSTM-Char-CRF	AR	FastText	63.1	48.5	54.9	60.8	46.8	52.8
	AR	MUSE	47.0	12.9	20.3	43.3	11.9	18.7
	EN		71.5	42.2	53.0	56.9	33.5	42.2
	EN+AR		68.1	47.7	56.1	64.6	45.3	53.2
	All		69.7	69.1	69.4	62.6	62.0	62.3
BERT-CRF	AR	BERT (multi)	66.7	73.2	69.8	63.7	69.9	66.7
	EN		26.4	66.7	37.8	21.6	54.6	30.9
	EN+AR		73.8	76.1	74.9	68.5	70.6	69.5
	All		71.9	74.5	73.2	67.7	70.2	68.9

Table 3: Comparison of performance **testing on Arabic** using different training modes comparing Bi-LSTM-Char-CRF with MUSE Embeddings to BERT-CRF