

## A Supplemental Material

### A.1 Representation Test

To evaluate the quality of learning representation, we design a mask experiment. We also adopt the (C)LSTM-(W)LSTM-CRF as the Baseline model.

This experiment aims to test the reasoning capacity of the bilateral model. We hide all the entity names in the test set and replaces the entity word with “UNK” token, i.e. “*Germany imported 47000 sheep from Britain*” becomes “*UNK imported 47000 sheep from UNK*” . As shown in Table 1, the augmented network achieves a better result than the baseline model, while the bilateral model forms a trade-off. This scenario is more subtle than the case of out-of-vocabulary (OOV) words because entities cannot differentiate each other from morphology. This model needs to infer the entity type based on the context information, which is a way to test the contextual representation. The result demonstrates that the augmented model can extract more features from context to infer the entity type.

Table 1: Results of mask test on the CoNLL-2003 English test set

Model	Recall	Precision	F1
Baseline	32.68	36.09	34.40
Augment	<b>41.36</b>	<b>41.51</b>	<b>41.43</b>
Bilateral	36.13	38.71	37.38

### A.2 Case Study

It is instructive to analyze the type and length of entities in the prediction results. We select some samples from the CoNLL-2003 English dataset. The Baseline model adopts the (C)LSTM-(W)LSTM-CRF model in Table ?? . As shown in Table 2, in sentence 1, the *Costa* is an ambiguity entity since the *Costa* represents a location in most cases but the context decides the *Costa* to be the person. This suggests that the baseline model focuses more on the word sense from the statistic level. The augmented model makes a correct prediction because it reduces the entity specificity by extracting more information from the context pattern. Sentence 4 demonstrates the same issue.

In sentence 2, the *English* is a “S-MISC” entity in most scenarios, but the *ENGLISH F.A. CUP* is a whole phrase here. This implies that the baseline model sometimes makes a greedy prediction about the current word. The augmented model enhanced the representations by considering more context information.

In sentence 3, from a syntax level we know the pattern “from {1} to {2}” means that the {1} and {2} are more likely in the same type. The baseline model predicted the “from ORG to LOC”. The augmented model fix the last pair to “from ORG to ORG” but missed the *NHL* entity in the other pair. The bilateral

Table 2: Examples on the CoNLL-2003 English test set where the blue and red label represent the correct and incorrect predictions respectively and [...]miss is the not recognized entity

Model	Sentence
Baseline	[Weah]PER has admitted head butting [Costa]LOC but said he reacted to racist taunts .
Augment	[Weah]PER has admitted head butting [Costa]PER but said he reacted to racist taunts .
Bilateral	[Weah]PER has admitted head butting [Costa]PER but said he reacted to racist taunts .
Baseline	SOCCER - [ENGLISH]MISC [F.A. CUP]MISC SECOND ROUND RESULT .
Augment	SOCCER - [ENGLISH F.A. CUP]MISC SECOND ROUND RESULT .
Bilateral	SOCCER - [ENGLISH F.A. CUP]MISC SECOND ROUND RESULT .
Baseline	( Corrects headline from [NBA]ORG to [NHL]ORG and corrects team name in second result from [La Clippers]ORG to [Ny Islanders]LOC .
Augment	( Corrects headline from [NBA]ORG to [NHL]miss and corrects team name in second result from [La Clippers]ORG to [Ny Islanders]ORG .
Bilateral	( Corrects headline from [NBA]ORG to [NHL]ORG and corrects team name in second result from [La Clippers]ORG to [Ny Islanders]ORG .
Baseline	High-flying Italy topped the league in a week of meagre returns on government bonds , [Salomon Brothers]PER said on Friday .
Augment	High-flying Italy topped the league in a week of meagre returns on government bonds , [Salomon Brothers]ORG said on Friday .
Bilateral	High-flying Italy topped the league in a week of meagre returns on government bonds , [Salomon Brothers]ORG said on Friday .

model made a correct prediction in both pairs. These results indicate that the bilateral model helps to improve the model robustness through enhancing the representations.