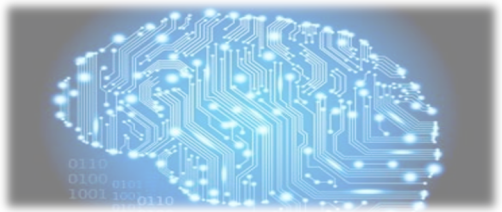
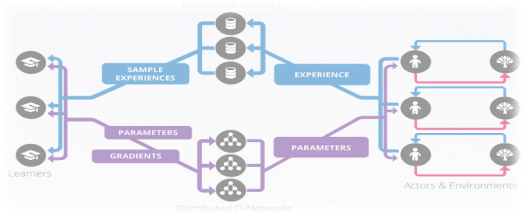

S2W 데이터셋 학습 실험 결과 보고 – 2 분석

20151322 이한용

S2W 데이터셋 학습 실험 결과 보고 – 2 분석

RAN -> MISC로 잘못 태그되는 문제

- train set 에서 멀웨어가 MISC으로 태깅되었고 랜섬웨어와 멀웨어가 문장에서 쓰이는 맥락이 비슷한데, 이 때문에 분류에 애매모호함이 생김
- 멀웨어가 랜섬웨어를 포함하므로 맥락적으로 MISC 태그가 RAN 태그를 포함함

- ['demand', 'ransom', 'Ransom', 'crypt', 'Crypt', 'payment', 'bitcoin', 'Bitcoin', 'lock', 'Lock’] 의 키워드가 token 주변 10 token 내의 범위에 나타났을 때 RAN으로 제대로 태깅된다고 가정해 봄
- 위와 같은 가정에서 성능 향상이 얼마나 나오는지 F1 score를 계산함

S2W 데이터셋 학습 실험 결과 보고 – 2 분석

이전 실험 결과

F1-scores	F1-score (micro)	F1-score (macro)	LOC	MISC	ORG	PER	RAN
실험1	0.8603	0.8540	0.9259	0.8823	0.8083	0.8148	0.8389
실험2	0.8600	0.8766	0.9516	0.8826	0.8242	0.9062	0.8184
실험3	0.8450	0.8609	0.9737	0.8798	0.8291	0.8406	0.7813
실험4	0.8989	0.8918	0.9296	0.9122	0.8590	0.8710	0.8872
실험5	0.8712	0.9065	0.9639	0.8742	0.8622	0.9730	0.8593
실험6	0.8543	0.8780	0.9620	0.8844	0.8204	0.9231	0.8004
실험7	0.9047	0.9096	0.9677	0.9011	0.8786	0.8889	0.9117
실험8	0.9198	0.8970	0.9388	0.9363	0.8418	0.8462	0.9220
실험9	0.8332	0.8554	0.9635	0.8452	0.8491	0.8276	0.7914
실험10	0.8764	0.8971	0.9811	0.8935	0.8418	0.9184	0.8508
평균 ± 분산	**0.8724 ± 0.0008**	**0.8827 ± 0.0004**	**0.9558 ± 0.0004**	**0.8892 ± 0.0006**	**0.8415 ± 0.0005**	**0.8810 ± 0.0025**	**0.8461 ± 0.0024**

S2W 데이터셋 학습 실험 결과 보고 – 2 분석

keyword를 잘 잡아냈다고 가정할 때 예상 결과

F1-scores	F1-score (micro)	F1-score (macro)	LOC	MISC	ORG	PER	RAN
실험1	0.9059	0.8832	0.9259	0.9124	0.8483	0.8148	0.9145
실험2	0.9081	0.9107	0.9516	0.9147	0.8412	0.9355	0.9104
실험3	0.8794	0.8852	0.9737	0.9008	0.8481	0.8529	0.8504
실험4	0.9361	0.9191	0.9296	0.9376	0.8787	0.9000	0.9495
실험5	0.9174	0.9303	0.9639	0.9069	0.8673	0.9730	0.9403
실험6	0.9230	0.9192	0.9620	0.9325	0.8524	0.9231	0.9262
실험7	0.9270	0.9227	0.9677	0.9134	0.8889	0.8889	0.9544
실험8	0.9296	0.9018	0.9388	0.9451	0.8418	0.8462	0.9374
실험9	0.8832	0.8838	0.9635	0.8815	0.8592	0.8276	0.8874
실험10	0.9234	0.9226	0.9811	0.9184	0.8468	0.9184	0.9484
평균 ± 분산	**0.9133 ± 0.0004**	**0.9079 ± 0.0003**	**0.9558 ± 0.0004**	**0.9163 ± 0.0003**	**0.8573 ± 0.0003**	**0.8880 ± 0.0026**	**0.9219 ± 0.0011**

S2W 데이터셋 학습 실험 결과 보고 – 2 분석

해결방법 제안

1. PER, ORG, LOC, MISC, RAN, MAL 의 6개 태그를 사용하여 MAL으로 태깅된 토큰에 대해 키워드가 주위에 있으면 RAN으로 태깅한다.
2. data augmentation 방법을 사용한다. (train set에 keyword가 들어간 문장을 더 추가한다.)
3. 다른 embedding을 추가한다. (BERT, ELMo, Pooled Flair, BytePairEmbeddings 등)
   - BERT나 ELMo, Pooled Flair 는 Contextualized embedding이므로 keyword에 대한 feature을 잘 잡아낼 것으로 예상함
   - 또한 일반적으로 Pooled Flair Embedding을 사용하면 모델의 성능이 좋아짐
   - 아래 표는 Pooled Flair Embedding을 설명한 논문에서 발췌함

	Approach	CoNLL-03 En	CoNLL-03 De	CoNLL-03 Nl	WNUT-17
(Pooled Flair Embedding)	Pooled Contextualized Embeddings$_{min}$	**93.18** ± 0.09	**88.27** ± 0.30	90.12 ± 0.14	49.07 ± 0.31
	Pooled Contextualized Embeddings$_{max}$	93.13 ± 0.09	88.05 ± 0.25	90.26 ± 0.10	49.05 ± 0.26
	Pooled Contextualized Embeddings$_{mean}$	93.10 ± 0.11	87.69 ± 0.27	**90.44** ± 0.20	**49.59** ± 0.41
(Flair Embedding)	Contextual String Emb. (Akbik et al., 2018)	92.86 ± 0.08	87.41 ± 0.13	90.16 ± 0.26	49.49 ± 0.75
	best published				
	BERT (Devlin et al., 2018)†	92.8			
	CVT+Multitask (Clark et al., 2018)†	92.6			
	ELMo (Peters et al., 2018)†	92.22			
	Stacked Multitask (Aguilar et al., 2018)†				45.55
	Character-LSTM (Lample et al., 2016)†	90.94	78.76	81.74	